PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, C07K 14/35, 16/12, C12Q 1/68, C12N 15/62, G01N 33/53

A2

(11) International Publication Number: **WO 98/16645**

(43) International Publication Date: 23 April 1998 (23.04.98)

(21) International Application Number: PCT/US97/18214

(22) International Filing Date: 7 October 1997 (07.10.97)

(30) Priority Data:

08/729,622	11 October 1996 (11.10.96)	US
08/818,111	13 March 1997 (13.03.97)	US

(71) Applicant: CORIXA CORPORATION [US/US]; 1124 Columbia Street, Seattle, WA 98104 (US).

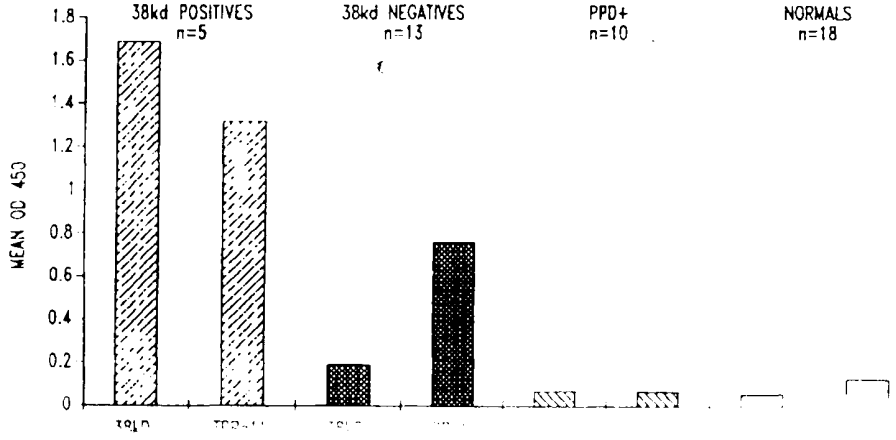
(72) Inventors: REED, Steven, G.; 2843 – 122nd Place N.E., Bellevue, WA 98005 (US). SKEIKY, Yasir, A., W.; 8327 – 25th Avenue N.W., Seattle, WA 98107 (US). DILLON, Davin, C.; 21607 N.E. 24th Street, Redmond, WA 98053 (US). CAMPOS-NETO, Antonio; 9308 Midship Court N.E., Bainbridge Island, WA 98021 (US). HOUGHTON, Raymond; 2636 – 242nd Place S.E., Bothell, WA 98021 (US). VEDVICK, Thomas, S.; 124 South 300th Place, Federal Way, WA 98003 (US). TWARDZIK, Daniel, R.; 10195 South Beach Drive, Bainbridge Island, WA 98110 (US). LODES, Michael, J.; 9223 – 36th Avenue S.W., Seattle, WA 98126 (US).

(74) Agents: MAKI, David, J. et al., Seed and Berry LLP, 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).

(81) Designated States: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel				
BY	Belarus						

COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

TECHNICAL FIELD

The present invention relates generally to the detection of *Mycobacterium tuberculosis* infection. The invention is more particularly related to polypeptides comprising a *Mycobacterium tuberculosis* antigen, or a portion or other variant thereof, and the use of such polypeptides for the serodiagnosis of *Mycobacterium tuberculosis* infection.

BACKGROUND OF THE INVENTION

Tuberculosis is a chronic, infectious disease, that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries, as well as an increasing problem in developed areas of the world, with about 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis will require effective vaccination and accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common Mycobacterium for this purpose is Bacillus Calmette Guerin (BCG).

site by 48-72 hours after injection, which indicates exposure to Mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of *M. tuberculosis* immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of CD4 T cells associated with human immunodeficiency virus (HIV) infection. Mycobacterium-reactive CD4 T cells have been shown to be potent producers of gamma-interferon (IFN-γ), which, in turn, has been shown to trigger the anti-mycobacterial effects of macrophages in mice. While the role of IFN-γ in humans is less clear, studies have shown that 1,25-dihydroxy-vitamin D3, either alone or in combination with IFN-γ or tumor necrosis factor-alpha, activates human macrophages to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN-γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection *see* Chan and Kaufmann, in *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved diagnostic methods for detecting tuberculosis. The present invention fulfills this need and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, the present invention provides compositions and methods for diagnosing tuberculosis. In one aspect, polypeptides are provided comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of such an antigen that differs only

(b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);

(c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 117);

(d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);

(e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);

(f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);

(g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);

(i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123);

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)

(k) Ala Gly Asp Thr Xaa Ile Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID NO: 130) or

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu Gly Gly Thr-Val Gln-Ala-Gly; (SEQ ID NO: 131)

wherein Xaa may be any amino acid.

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132) or

(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124)

wherein Xaa may be any amino acid.

In another embodiment, the soluble *M. tuberculosis* antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

In a related aspect, the polypeptides comprise an antigenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, wherein the antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196 or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides, recombinant expression vectors comprising these DNA sequences and host cells transformed or transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

In further aspects of the subject invention, methods and diagnostic kits are provided for detecting tuberculosis in a patient. The methods comprise: (a) contacting a [illegible] cerebrospinal fluid and urine. The diagnostic kit [illegible]

The present invention also provides methods for detecting *M. tuberculosis* infection comprising: (a) obtaining a biological sample from a patient; (b) contacting the sample with at least one oligonucleotide primer in a polymerase chain reaction, the oligonucleotide primer being specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers. In one embodiment, the oligonucleotide primer comprises at least about 10 contiguous nucleotides of such a DNA sequence.

In a further aspect, the present invention provides a method for detecting *M. tuberculosis* infection in a patient comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with an oligonucleotide probe specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe. In one embodiment, the oligonucleotide probe comprises at least about 15 contiguous nucleotides of such a DNA sequence.

In yet another aspect, the present invention provides antibodies, both polyclonal and monoclonal, that bind to the polypeptides described above, as well as methods for their use in the detection of *M. tuberculosis* infection.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1A and B illustrate the stimulation of proliferation and interferon γ production in T cells derived from a first and a second *M. tuberculosis*-immune donor, respectively, by the 14 Kd, 20 Kd and 26 Kd antigens described in Example 1.

Figure 3A illustrates the stimulation of proliferation in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, recombinant TbH-9 and a control antigen, TbRa11.

Figure 3B illustrates the stimulation of interferon-γ production in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, PPD and recombinant TbH-9.

Figure 4 illustrates the reactivity of two representative polypeptides with sera from *M tuberculosis*-infected and uninfected individuals, as compared to the reactivity of bacterial lysate.

Figure 5 shows the reactivity of four representative polypeptides with sera from *M. tuberculosis*-infected and uninfected individuals, as compared to the reactivity of the 38 kD antigen.

Figure 6 shows the reactivity of recombinant 38 kD and TbRa11 antigens with sera from *M. tuberculosis* patients, PPD positive donors and normal donors.

Figure 7 shows the reactivity of the antigen TbRa2A with 38 kD negative sera.

Figure 8 shows the reactivity of the antigen of SEQ ID NO: 60 with sera from *M. tuberculosis* patients and normal donors.

Figure 9 illustrates the reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with sera from *M tuberculosis* patients, PPD positive donors and normal donors as determined by indirect ELISA

Figure 10 illustrates the reactivity of the recombinant antigen TbH-33 (SEQ ID NO: 140) with sera from *M. tuberculosis* patients and from normal donors, and with a pool of sera from *M. tuberculosis* patients, as determined both by direct and indirect ELISA

Figure 11 illustrates the reactivity of increasing concentrations of the recombinant antigen TbH-33 (SEQ ID NO: 140) with sera from M. tuberculosis patients and from normal donors as determined by ELISA.

SEQ. ID NO. 5 is the DNA sequence of TbRa13.

SEQ. ID NO. 6 is the DNA sequence of TbRa16.

SEQ. ID NO. 7 is the DNA sequence of TbRa17.

SEQ. ID NO. 8 is the DNA sequence of TbRa18.

SEQ. ID NO. 9 is the DNA sequence of TbRa19.

SEQ. ID NO. 10 is the DNA sequence of TbRa24.

SEQ. ID NO. 11 is the DNA sequence of TbRa26.

SEQ. ID NO. 12 is the DNA sequence of TbRa28.

SEQ. ID NO. 13 is the DNA sequence of TbRa29.

SEQ. ID NO. 14 is the DNA sequence of TbRa2A.

SEQ. ID NO. 15 is the DNA sequence of TbRa3.

SEQ. ID NO. 16 is the DNA sequence of TbRa32.

SEQ. ID NO. 17 is the DNA sequence of TbRa35.

SEQ. ID NO. 18 is the DNA sequence of TbRa36.

SEQ. ID NO. 19 is the DNA sequence of TbRa4.

SEQ. ID NO. 20 is the DNA sequence of TbRa9.

SEQ. ID NO. 21 is the DNA sequence of TbRaB.

SEQ. ID NO. 22 is the DNA sequence of TbRaC.

SEQ. ID NO. 23 is the DNA sequence of TbRaD

SEQ. ID NO. 24 is the DNA sequence of YYWCPG

SEQ. ID NO. 25 is the DNA sequence of AAMK.

SEQ. ID NO. 26 is the DNA sequence of TbL-23.

SEQ. ID NO. 27 is the DNA sequence of TbL-24.

SEQ. ID NO. 28 is the DNA sequence of TbL-25

SEQ. ID NO. 29 is the DNA sequence of TbL-28.

SEQ. ID NO. 35 is the DNA sequence of TbM-3.

SEQ. ID NO. 36 is the DNA sequence of TbM-6.

SEQ. ID NO. 37 is the DNA sequence of TbM-7.

SEQ. ID NO. 38 is the DNA sequence of TbM-9.

SEQ. ID NO. 39 is the DNA sequence of TbM-12.

SEQ. ID NO. 40 is the DNA sequence of TbM-13

SEQ. ID NO. 41 is the DNA sequence of TbM-14.

SEQ. ID NO. 42 is the DNA sequence of TbM-15.

SEQ. ID NO. 43 is the DNA sequence of TbH-4.

SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.

SEQ. ID NO. 45 is the DNA sequence of TbH-12.

SEQ. ID NO. 46 is the DNA sequence of Tb38-1.

SEQ. ID NO. 47 is the DNA sequence of Tb38-4.

SEQ. ID NO. 48 is the DNA sequence of TbL-17.

SEQ. ID NO. 49 is the DNA sequence of TbL-20.

SEQ. ID NO. 50 is the DNA sequence of TbL-21.

SEQ. ID NO. 51 is the DNA sequence of TbH-16.

SEQ. ID NO. 52 is the DNA sequence of DPEP.

SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.

SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.

SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.

SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.

SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.

SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.

SEQ. ID NO. 59 is the protein sequence of AFES N-terminal Antigen.

SEQ. ID NO. [illegible] is the [illegible] of TbM-3 [illegible]

SEQ. ID NO. [illegible]

SEQ. ID NO. 65 is the deduced amino acid sequence of TbRa10.

SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa11.

SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa12.

SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa13.

SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa16.

SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa17.

SEQ. ID NO. 71 is the deduced amino acid sequence of TbRa18.

SEQ. ID NO. 72 is the deduced amino acid sequence of TbRa19.

SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa24.

SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa26.

SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa28.

SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa29.

SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa2A.

SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa3.

SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa32.

SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa35.

SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa36.

SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa4.

SEQ. ID NO. 83 is the deduced amino acid sequence of TbRa9.

SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaB.

SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaC.

SEQ. ID NO. 86 is the deduced amino acid sequence of TbRaD.

SEQ. ID NO. 87 is the deduced amino acid sequence of YYWCPG.

SEQ. ID NO. 88 is the deduced amino acid sequence of TbAAMK.

SEQ. ID NO. 89 is the deduced amino acid sequence of Tb38-1

SEQ. ID NO. [illegible]

SEQ. ID NO. 95 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 96 is the DNA sequence of DPV.

SEQ. ID NO. 97 is the deduced amino acid sequence of DPV.

SEQ. ID NO. 98 is the DNA sequence of ESAT-6.

SEQ. ID NO. 99 is the deduced amino acid sequence of ESAT-6.

SEQ. ID NO. 100 is the DNA sequence of TbH-8-2.

SEQ. ID NO. 101 is the DNA sequence of TbH-9FL.

SEQ. ID NO. 102 is the deduced amino acid sequence of TbH-9FL.

SEQ. ID NO. 103 is the DNA sequence of TbH-9-1.

SEQ. ID NO. 104 is the deduced amino acid sequence of TbH-9-1.

SEQ. ID NO. 105 is the DNA sequence of TbH-9-4.

SEQ. ID NO. 106 is the deduced amino acid sequence of TbH-9-4.

SEQ. ID NO. 107 is the DNA sequence of Tb38-1F2 IN.

SEQ. ID NO. 108 is the DNA sequence of Tb38-1F2 RP.

SEQ. ID NO. 109 is the deduced amino acid sequence of Tb37-FL.

SEQ. ID NO. 110 is the deduced amino acid sequence of Tb38-IN.

SEQ. ID NO. 111 is the DNA sequence of Tb38-1F3.

SEQ. ID NO. 112 is the deduced amino acid sequence of Tb38-1F3.

SEQ. ID NO. 113 is the DNA sequence of Tb38-1F5.

SEQ. ID NO. 114 is the DNA sequence of Tb38-1F6.

SEQ. ID NO. 115 is the deduced N-terminal amino acid sequence of DPV.

SEQ. ID NO. 116 is the deduced N-terminal amino acid sequence of AVGS.

SEQ. ID NO. 117 is the deduced N-terminal amino acid sequence of AAMK.

SEQ. ID NO. 118 is the deduced N-terminal amino acid sequence of YYWC.

SEQ. ID NO. 119 is the deduced N-terminal amino acid sequence of DIGS.

SEQ. ID NO. [illegible]

SEQ. ID NO. [illegible] deduced [illegible] amino acid sequence of DPAS

SEQ. ID NO. [illegible] DPPV [illegible]

SEQ ID NO. 125-128 are the protein sequences of four DPPD cyanogen bromide fragments.

SEQ ID NO. 129 is the N-terminal protein sequence of XDS antigen.

SEQ ID NO. 130 is the N-terminal protein sequence of AGD antigen.

SEQ ID NO. 131 is the N-terminal protein sequence of APE antigen.

SEQ ID NO. 132 is the N-terminal protein sequence of XYI antigen.

SEQ ID NO. 133 is the DNA sequence of TbH-29.

SEQ ID NO. 134 is the DNA sequence of TbH-30.

SEQ ID NO. 135 is the DNA sequence of TbH-32.

SEQ ID NO. 136 is the DNA sequence of TbH-33.

SEQ ID NO. 137 is the predicted amino acid sequence of TbH-29.

SEQ ID NO. 138 is the predicted amino acid sequence of TbH-30.

SEQ ID NO. 139 is the predicted amino acid sequence of TbH-32.

SEQ ID NO. 140 is the predicted amino acid sequence of TbH-33.

SEQ ID NO: 141-146 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 147 is the DNA sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 148 is the amino acid sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 149 is the DNA sequence of the M. tuberculosis antigen 38 kD.

SEQ ID NO: 150 is the amino acid sequence of the M. tuberculosis antigen 38 kD.

SEQ ID NO: 151 is the DNA sequence of XP14.

SEQ ID NO: 152 is the DNA sequence of XP24.

SEQ ID NO: 153 is the DNA sequence of XP31.

SEQ ID NO: 158 is the DNA sequence of XP27.

SEQ ID NO: 159 is the DNA sequence of XP36.

SEQ ID NO: 160 is the 5' DNA sequence of XP4.

SEQ ID NO: 161 is the 5' DNA sequence of XP5.

SEQ ID NO: 162 is the 5' DNA sequence of XP17.

SEQ ID NO: 163 is the 5' DNA sequence of XP30

SEQ ID NO: 164 is the 5' DNA sequence of XP2.

SEQ ID NO: 165 is the 3' DNA sequence of XP2.

SEQ ID NO: 166 is the 5' DNA sequence of XP3.

SEQ ID NO: 167 is the 3' DNA sequence of XP3.

SEQ ID NO: 168 is the 5' DNA sequence of XP6.

SEQ ID NO: 169 is the 3' DNA sequence of XP6.

SEQ ID NO: 170 is the 5' DNA sequence of XP18.

SEQ ID NO: 171 is the 3' DNA sequence of XP18.

SEQ ID NO: 172 is the 5' DNA sequence of XP19.

SEQ ID NO: 173 is the 3' DNA sequence of XP19.

SEQ ID NO: 174 is the 5' DNA sequence of XP22.

SEQ ID NO: 175 is the 3' DNA sequence of XP22.

SEQ ID NO: 176 is the 5' DNA sequence of XP25

SEQ ID NO: 177 is the 3' DNA sequence of XP25.

SEQ ID NO: 178 is the full-length DNA sequence of TbH4-XP1.

SEQ ID NO: 179 is the predicted amino acid sequence of TbH4-XP1.

SEQ ID NO: 180 is the predicted amino acid sequence encoded by the reverse complement of TbH4-XP1.

SEQ ID NO: 181 is a first predicted amino acid sequence encoded by XP36.

SEQ ID NO: 182 is [illegible]

[illegible]

SEQ ID NO: [illegible]

SEQ ID NO: 186 is the DNA sequence of RDIF8.

SEQ ID NO: 187 is the DNA sequence of RDIF10.

SEQ ID NO: 188 is the DNA sequence of RDIF11.

SEQ ID NO: 189 is the predicted amino acid sequence of RDIF2.

SEQ ID NO: 190 is the predicted amino acid sequence of RDIF5.

SEQ ID NO: 191 is the predicted amino acid sequence of RDIF8.

SEQ ID NO: 192 is the predicted amino acid sequence of RDIF10.

SEQ ID NO: 193 is the predicted amino acid sequence of RDIF11.

SEQ ID NO: 194 is the 5' DNA sequence of RDIF12.

SEQ ID NO: 195 is the 3' DNA sequence of RDIF12.

SEQ ID NO: 196 is the DNA sequence of RDIF7.

SEQ ID NO: 197 is the predicted amino acid sequence of RDIF7.

SEQ ID NO: 198 is the DNA sequence of DIF2-1.

SEQ ID NO: 199 is the predicted amino acid sequence of DIF2-1.

SEQ ID NO: 200-207 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD, Tb38-1 and DPEP (hereinafter referred to as TbF-2).

SEQ ID NO: 208 is the DNA sequence of the fusion protein TbF-2.

SEQ ID NO: 209 is the amino acid sequence of the fusion protein TbF-2.

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to compositions and methods for diagnosing tuberculosis. The compositions of the subject invention include polypeptides that comprise at least one antigenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and or modifications

a polypeptide comprising an antigenic portion of one of the above antigens may consist entirely of the antigenic portion, or may contain additional sequences. The additional sequences may be derived from the native *M. tuberculosis* antigen or may be heterologous, and such sequences may (but need not) be antigenic.

An "antigenic portion" of an antigen (which may or may not be soluble) is a portion that is capable of reacting with sera obtained from an *M. tuberculosis*-infected individual (*i.e.*, generates an absorbance reading with sera from infected individuals that is at least three standard deviations above the absorbance obtained with sera from uninfected individuals, in a representative ELISA assay described herein). An "*M. tuberculosis*-infected individual" is a human who has been infected with *M. tuberculosis* (*e.g.*, has an intradermal skin test response to PPD that is at least 0.5 cm in diameter). Infected individuals may display symptoms of tuberculosis or may be free of disease symptoms. Polypeptides comprising at least an antigenic portion of one or more *M. tuberculosis* antigens as described herein may generally be used, alone or in combination, to detect tuberculosis in a patient.

The compositions and methods of this invention also encompass variants of the above polypeptides. A "variant," as used herein, is a polypeptide that differs from the native antigen only in conservative substitutions and/or modifications, such that the antigenic properties of the polypeptide are retained. Such variants may generally be identified by modifying one of the above polypeptide sequences, and evaluating the antigenic properties of the modified polypeptide using, for example, the representative procedures described herein.

A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val

translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

In a related aspect, combination polypeptides are disclosed. A "combination polypeptide" is a polypeptide comprising at least one of the above antigenic portions and one or more additional antigenic *M. tuberculosis* sequences, which are joined via a peptide linkage into a single amino acid chain. The sequences may be joined directly (*i.e.*, with no intervening amino acids) or may be joined by way of a linker sequence (*e.g.*, Gly-Cys-Gly) that does not significantly diminish the antigenic properties of the component polypeptides.

In general, *M. tuberculosis* antigens, and DNA sequences encoding such antigens, may be prepared using any of a variety of procedures. For example, soluble antigens may be isolated from *M. tuberculosis* culture filtrate by procedures known to those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens may then be evaluated for a desired property, such as the ability to react with sera obtained from an *M. tuberculosis*-infected individual. Such screens may be performed using the representative methods described herein. Antigens may then be partially sequenced using, for example, traditional Edman chemistry. See Edman and Berg, *Eur. J. Biochem.* 80:116-132, 1967.

Antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and expressed in an appropriate host. DNA molecules encoding soluble antigens may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (*e.g.*, rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may

[illegible] Sprin[illegible]

Harbor Laboratory [illegible]

DNA sequences encoding soluble antigens may also be obtained by screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as described (for example) in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY (and references cited therein). Polymerase chain reaction (PCR) may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

Regardless of the method of preparation, the antigens described herein are "antigenic." More specifically, the antigens have the ability to react with sera obtained from an *M. tuberculosis*-infected individual. Reactivity may be evaluated using, for example, the representative ELISA assays described herein, where an absorbance reading with sera from infected individuals that is at least three standard deviations above the absorbance obtained with sera from uninfected individuals is considered positive.

Antigenic portions of *M. tuberculosis* antigens may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited therein. Such techniques include screening polypeptide portions of the native antigen for antigenic properties. The representative ELISAs described herein may generally be employed in these screens. An antigenic portion of a polypeptide is a portion that, within such representative assays, generates a signal in such assays that is substantially similar to that generated by the full length antigen. In other words, an antigenic portion of a *M. tuberculosis* antigen generates at least about 20%, and preferably about 100%, of the signal induced by the full length antigen

commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc. 85*:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides as described herein. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line, such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the

In certain specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen), where the antigen has one of the following N-terminal sequences:

(a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115),

(b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);

(c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 117);

(d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);

(e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);

(f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);

(g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);

(i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123);

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp;

amino acid sequence of which is provided in SEQ ID NO: 53. A DNA sequence encoding the antigen identified as (a) above is provided in SEQ ID NO: 96; its deduced amino acid sequence is provided in SEQ ID NO: 97. A DNA sequence corresponding to antigen (d) above is provided in SEQ ID NO: 24, a DNA sequence corresponding to antigen (c) is provided in SEQ ID NO: 25 and a DNA sequence corresponding to antigen (I) is disclosed in SEQ ID NO: 94 and its deduced amino acid sequence is provided in SEQ ID NO: 95.

In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132) or

(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124)

wherein Xaa may be any amino acid, preferably a cysteine residue.

In other specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, (b) the complements of such DNA sequences, or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In further specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen), which may or may not be soluble, that comprises one or more of the amino acid

In [illegible] embodiments discussed above, the *M. tuberculosis* antigens

or more of DNA sequences specifically recited herein. "Substantial homology," as used herein, refers to DNA sequences that are capable of hybridizing under moderately stringent conditions. Suitable moderately stringent conditions include prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5X SSC, overnight or, in the event of cross-species homology, at 45°C with 0.5X SSC; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS). Such hybridizing DNA sequences are also within the scope of this invention, as are nucleotide sequences that, due to code degeneracy, encode an immunogenic polypeptide that is encoded by a hybridizing DNA sequence.

In a related aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen described above or ESAT-6 (SEQ ID NOS: 98 and 99), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptides by a distance sufficient to ensure that each polypeptide folds into its

or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene 40*:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA 83*:8258-8562, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric hindrance.

In another aspect, the present invention provides methods for using the polypeptides described above to diagnose tuberculosis. In this aspect, methods are provided for detecting *M. tuberculosis* infection in a biological sample, using one or more of the above polypeptides, alone or in combination. In embodiments in which multiple polypeptides are employed, polypeptides other than those specifically described herein, such as the 38 kD antigen described in Andersen and Hansen, *Infect. Immun. 57*:2481-2488, 1989, may be included. As used herein, a "biological sample" is any antibody-containing sample obtained from a patient. Preferably, the sample is whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid or urine. More preferably, the sample is a blood, serum or plasma sample obtained from a patient or a blood supply. The polypeptide(s) are used in an assay, as described below, to determine the presence or absence of antibodies to the polypeptide(s) in the sample, relative to a predetermined cut off value. The presence of such antibodies indicates previous sensitization to mycobacterial antigens which may be indicative of tuberculosis.

In embodiments in which more than one polypeptide is employed, the polypeptides used are preferably complementary (*i.e.*, one component polypeptide will tend

formulated that are capable of detecting infection in most, or all, of the samples tested. Such polypeptides are complementary. For example, approximately 25-30% of sera from tuberculosis-infected individuals are negative for antibodies to any single protein, such as the 38 kD antigen mentioned above. Complementary polypeptides may, therefore, be used in combination with the 38 kD antigen to improve sensitivity of a diagnostic test.

There are a variety of assay formats known to those of ordinary skill in the art for using one or more polypeptides to detect antibodies in a sample. *See*, *e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988, which is incorporated herein by reference. In a preferred embodiment, the assay involves the use of polypeptide immobilized on a solid support to bind to and remove the antibody from the sample. The bound antibody may then be detected using a detection reagent that contains a reporter group. Suitable detection reagents include antibodies that bind to the antibody/polypeptide complex and free polypeptide labeled with a reporter group (*e.g.*, in a semi-competitive assay). Alternatively, a competitive assay may be utilized, in which an antibody that binds to the polypeptide is labeled with a reporter group and allowed to bind to the immobilized antigen after incubation of the antigen with the sample. The extent to which components of the sample inhibit the binding of the labeled antibody to the polypeptide is indicative of the reactivity of the sample with the immobilized polypeptide.

The solid support may be any solid material known to those of ordinary skill in the art to which the antigen may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The polypeptides may be bound to the solid support using a variety of

membrane is preferred. In such cases, adsorption may be achieved by contacting the polypeptide, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of polypeptide ranging from about 10 ng to about 1 μg, and preferably about 100 ng, is sufficient to bind an adequate amount of antigen.

Covalent attachment of polypeptide to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the polypeptide. For example, the polypeptide may be bound to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the polypeptide (*see, e.g.*, Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is an enzyme linked immunosorbent assay (ELISA). This assay may be performed by first contacting a polypeptide antigen that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that antibodies to the polypeptide within the sample are allowed to bind to the immobilized polypeptide. Unbound sample is then removed from the immobilized polypeptide and a detection reagent capable of binding to the immobilized antibody-polypeptide complex is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific detection reagent.

More specifically, once the polypeptide is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO) may be employed. The

[illegible] period [illegible] to detect the presence of antibody-antigen

of binding that is at least 95% of that achieved at equilibrium between bound and unbound antibody. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. Detection reagent may then be added to the solid support. An appropriate detection reagent is any compound that binds to the immobilized antibody-polypeptide complex and that can be detected by any of a variety of means known to those in the art. Preferably, the detection reagent contains a binding agent (such as, for example, Protein A, Protein G, immunoglobulin, lectin or free antigen) conjugated to a reporter group. Preferred reporter groups include enzymes (such as horseradish peroxidase), substrates, cofactors, inhibitors, dyes, radionuclides, luminescent groups, fluorescent groups and biotin. The conjugation of binding agent to reporter group may be achieved using standard methods known to those of ordinary skill in the art. Common binding agents may also be purchased conjugated to a variety of reporter groups from many commercial sources (*e.g.*, Zymed Laboratories, San Francisco, CA, and Pierce, Rockford, IL).

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound antibody. An appropriate amount of time may generally be determined from the manufacturer's instructions or by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods

To determine the presence or absence of anti-*M. tuberculosis* antibodies in the sample, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value is the average mean signal obtained when the immobilized antigen is incubated with samples from an uninfected patient. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for tuberculosis. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology. A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, pp. 106-107. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for tuberculosis.

In a related embodiment, the assay is performed in a rapid flow-through or strip test format, wherein the antigen is immobilized on a membrane, such as nitrocellulose. In the flow-through test, antibodies within the sample bind to the immobilized polypeptide as the sample passes through the membrane. A detection reagent (*e.g.*, protein A-colloidal gold) then binds to the antibody polypeptide complex as the solution containing the detection reagent flows through the membrane. The detection of bound detection reagent may then be

[illegible] of bound detection reagent at the polypeptide indicates the

detection reagent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of polypeptide immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of antibodies that would be sufficient to generate a positive signal in an ELISA, as discussed above. Preferably, the amount of polypeptide immobilized on the membrane ranges from about 25 ng to about 1 μg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount (*e.g.*, one drop) of patient serum or blood.

Of course, numerous other assay protocols exist that are suitable for use with the polypeptides of the present invention. The above descriptions are intended to be exemplary only.

In yet another aspect, the present invention provides antibodies to the inventive polypeptides. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one such technique, an immunogen comprising the antigenic polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep and goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

Antibodies may be used in diagnostic tests to detect the presence of *M tuberculosis* antigens using assays similar to those detailed above and other techniques well known to those of skill in the art, thereby providing a method for detecting *M tuberculosis* infection in a patient.

Diagnostic reagents of the present invention may also comprise DNA sequences encoding one or more of the above polypeptides, or one or more portions thereof. For example, at least two oligonucleotide primers may be employed in a polymerase chain

[illegible] cDNA [illegible] techniques well known in the art, such as gel electrophoresis

present invention may be used in a hybridization assay to detect the presence of an inventive polypeptide in a biological sample.

As used herein, the term "oligonucleotide primer/probe specific for a DNA molecule" means an oligonucleotide sequence that has at least about 80%, preferably at least about 90% and more preferably at least about 95%, identity to the DNA molecule in question. Oligonucleotide primers and/or probes which may be usefully employed in the inventive diagnostic methods preferably have at least about 10-40 nucleotides. In a preferred embodiment, the oligonucleotide primers comprise at least about 10 contiguous nucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Preferably, oligonucleotide probes for use in the inventive diagnostic methods comprise at least about 15 contiguous oligonucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Techniques for both PCR based assays and hybridization assays are well known in the art (see, for example, Mullis *et al. Ibid*; Ehrlich, *Ibid*). Primers or probes may thus be used to detect *M. tuberculosis*-specific sequences in biological samples. DNA probes or primers comprising oligonucleotide sequences described above may be used alone, in combination with each other, or with previously identified sequences, such as the 38 kD antigen discussed above.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

EXAMPLE 1

PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES

[illegible] of polypeptides as described below [illegible]

M. tuberculosis (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45 μ filter into a sterile 2.5 L bottle. The media was then filtered through a 0.2 μ filter into a sterile 4 L bottle. NaN_3 was then added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold room.

The culture filtrate was concentrated by placing the filtrate in a 12 L reservoir that had been autoclaved and feeding the filtrate into a 400 ml Amicon stir cell which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

The culture filtrate was then dialyzed into 0.1% ammonium bicarbonate using a 8,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were then dialyzed against 0.01 mM 1,3 bis[tris(hydroxymethyl)-methylamino]propane, pH 7.5 (Bis-Tris propane buffer), the initial conditions for anion exchange chromatography. Fractionation was performed using gel profusion chromatography on a POROS 146 II Q/M anion exchange column 4.6 mm x 100 mm (Perseptive BioSystems, Framingham, MA) equilibrated in 0.01 mM Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

The pools of polypeptides eluting from the ion exchange column were dialyzed against distilled water and lyophilized. The resulting material was dissolved in 0.1% trifluoroacetic acid (TFA)/40% [illegible] and fractionated by HPLC

to maximize the purity of the individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T cells were shown to proliferate in response to PPD and crude soluble proteins from MTB were cultured in medium comprising RPMI 1640 supplemented with 10% pooled human serum and 50 μg/ml gentamicin. Purified polypeptides were added in duplicate at concentrations of 0.5 to 10 μg/ml. After six days of culture in 96-well round-bottom plates in a volume of 200 μl, 50 μl of medium was removed from each well for determination of IFN-γ levels, as described below. The plates were then pulsed with 1 μCi/well of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

IFN-γ was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to human IFN-γ (Chemicon) in PBS for four hours at room temperature. Wells were then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates were then washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates were incubated overnight at room temperature. The plates were again washed and a polyclonal rabbit anti-human IFN-γ serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Jackson Labs.) was added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added. The reaction was

For sequencing, the polypeptides were individually dried onto Biobrene™ (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide by comparing the retention time of the PTH amino acid derivative to the appropriate PTH derivative standards.

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

(a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Xaa-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 54);

(b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 55);

(c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 56);

(d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 57);

(e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 58);

(f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 59);

(g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ala-Pro-Pro-Ala (SEQ ID NO: 60); and

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 61);

[illegible] purification step previously described was [illegible]

City, CA) with a 7 micron pore size, column size 1 mm x 100 mm, in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1%/minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80 µl/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller components and a polypeptide was obtained which was shown to have a molecular weight of 12,054 Kd (by mass spectrometry) and the following N-terminal sequence:

(i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Asp-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-Ala-Asp (SEQ ID NO: 62).

This polypeptide was shown to induce proliferation and IFN-γ production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above. Following dialysis against Bis-Tris propane buffer, at pH 5.5, fractionation was performed using anion exchange chromatography on a Poros QE column 4.6 x 100 mm (Perseptive Biosystems) equilibrated in Bis-Tris propane buffer pH 5.5. Polypeptides were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and resuspended

The fraction with biological activity was separated into one major peak plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd, 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequences, respectively:

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID NO: 130) and

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131), wherein Xaa may be any amino acid.

Using the assays described above, these polypeptides were shown to induce proliferation and IFN-γ production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a),.(c), (d) and (g) above were obtained by screening a *M. tuberculosis* genomic library using ^{32}P end labeled degenerate oligonucleotides corresponding to the N-terminal sequence and containing *M. tuberculosis* codon bias. The screen performed using a probe corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID NO: 96. The polypeptide encoded by SEQ ID NO: 96 is provided in SEQ ID NO: 97. The screen performed using a probe corresponding to antigen (g) above identified a clone having the sequence provided in SEQ ID NO: 52. The polypeptide encoded by SEQ ID NO: 52 is provided in SEQ ID NO: 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID NO: 24, and the screen performed with a probe corresponding to antigen (c) identified a clone having the sequence provided in SEQ ID NO: 25

The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen an *M. tuberculosis* library and a full length copy of the *M. tuberculosis* homologue was obtained (SEQ ID NO: 94).

The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

In the proliferation and IFN-γ assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1:

TABLE 1

RESULTS OF PBMC PROLIFERATION AND IFN-γ ASSAYS

Sequence	Proliferation	IFN-γ
(a)	+	-
(c)	+++	+++
(d)	++	++
(g)	+++	+++
(h)	+++	+++

In Table 1, responses that gave a stimulation index (SI) of between 2 and 4

indicate that these antigens are capable of inducing proliferation and/or interferon-γ production.

EXAMPLE 2

USE OF PATIENT SERA TO ISOLATE *M. TUBERCULOSIS* ANTIGENS

This example illustrates the isolation of antigens from *M. tuberculosis* lysate by screening with serum from *M. tuberculosis*-infected individuals.

Dessicated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with 20 mM Tris pH 7.5 and bound proteins eluted with 1M NaCl. The NaCl elute was dialyzed overnight against 10 mM Tris, pH 7.5. Dialyzed solution was treated with DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with α-D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10 (Amicon, Beverley, MA) and screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any

degenerate oligonucleotides corresponding to the N-terminal sequence of SEQ ID NO:137. A clone was identified having the DNA sequence provided in SEQ ID NO: 198. This sequence was found to encode the amino acid sequence provided in SEQ ID NO: 199. Comparison of these sequences with those in the genebank revealed some similarity to sequences previously identified in *M. tuberculosis* and *M. bovis*.

EXAMPLE 3

PREPARATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

This example illustrates the preparation of DNA sequences encoding *M. tuberculosis* antigens by screening a *M. tuberculosis* expression library with sera obtained from patients infected with *M. tuberculosis*, or with anti-sera raised against *M. tuberculosis* antigens.

A. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST M. TUBERCULOSIS SUPERNATANT

Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The DNA was randomly sheared and used to construct an expression library using the Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and Erdman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis* cultures. Specifically, the rabbit was first immunized subcutaneously with 200 µg of protein antigen in a total volume of 2 ml containing 100 µg muramyl dipeptide (Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's adjuvant. Finally, the rabbit was immunized intravenously

Thirty two clones were purified. Of these, 25 represent sequences that have not been previously identified in *M. tuberculosis*. Proteins were induced by IPTG and purified by gel elution, as described in Skeiky et al., *J. Exp. Med. 181*:1527-1537, 1995. Representative partial sequences of DNA molecules identified in this screen are provided in SEQ ID NOS: 1-25. The corresponding predicted amino acid sequences are shown in SEQ ID NOS: 64-88.

On comparison of these sequences with known sequences in the gene bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID NOS: 77, 69, 71, 76) show some homology to sequences previously identified in *Mycobacterium leprae* but not in *M. tuberculosis*. TbRA11, TbRA26, TbRA28 and TbDPEP (SEQ ID NOS: 66, 74, 75, 53) have been previously identified in *M. tuberculosis*. No significant homologies were found to TbRA1, TbRA3, TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRA19, TbRA29, TbRA32, TbRA36 and the overlapping clones TbRA35 and TbRA12 (SEQ ID NOS: 64, 78, 82, 83, 65, 68, 76, 72, 76, 79, 81, 80, 67, respectively). The clone TbRa24 is overlapping with clone TbRa29.

B. USE OF SERA FROM PATIENTS HAVING PULMONARY OR PLEURAL TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

The genomic DNA library described above, and an additional H37Rv library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial Sau3A digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Ca). Three different pools of sera, each containing sera obtained from three individuals with active pulmonary or pleural disease, were used in

lacked increased reactivity with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

Thirty two clones were purified. Of these, 31 represented sequences that had not been previously identified in human *M. tuberculosis*. Representative sequences of the DNA molecules identified are provided in SEQ ID NOS:: 26-51 and 100. Of these, TbH-8-2 (SEQ. ID NO. 100) is a partial clone of TbH-8, and TbH-4 (SEQ. ID NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1, TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID NOS.: 89-93. Comparison of these sequences with known sequences in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infect. Immun.* 63:1710-1717, 1995).

Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2, Tb38-1F3, Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS: 107, 108, 111, 113, and 114). (SEQ ID NOS: 107 and 108 are non-contiguous sequences from clone Tb38-1F2.) Two open reading frames

[illegible] H37Rv library. TbH-9-F1 (SEQ. ID NO. [illegible], which

ID NO. 105) is a partial clone of TbH-8. The deduced amino acid sequences for these three clones are presented in SEQ ID NOS: 102, 104 and 106.

Further screening of the *M. tuberculosis* genomic DNA library, as described above, resulted in the recovery of ten additional reactive clones, representing seven different genes. One of these genes was identified as the 38 Kd antigen discussed above, one was determined to be identical to the 14Kd alpha crystallin heat shock protein previously shown to be present in *M. tuberculosis*, and a third was determined to be identical to the antigen TbH-8 described above. The determined DNA sequences for the remaining five clones (hereinafter referred to as TbH-29, TbH-30, TbH-32 and TbH-33) are provided in SEQ ID NO: 133-136, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 137-140, respectively. The DNA and amino acid sequences for these antigens were compared with those in the gene bank as described above. No homologies were found to the 5' end of TbH-29 (which contains the reactive open reading frame), although the 3' end of TbH-29 was found to be identical to the *M. tuberculosis* cosmid Y227. TbH-32 and TbH-33 were found to be identical to the previously identified *M. tuberculosis* insertion element IS6110 and to the *M. tuberculosis* cosmid Y50, respectively. No significant homologies to TbH-30 were found.

Positive phagemid from this additional screening were used to infect *E. coli* XL-1 Blue MRF', as described in Sambrook et al., *supra*. Induction of recombinant protein was accomplished by the addition of IPTG. Induced and uninduced lysates were run in duplicate on SDS-PAGE and transferred to nitrocellulose filters. Filters were reacted with human *M. tuberculosis* sera (1:200 dilution) reactive with TbH and a rabbit sera (1:200 or 1:250 dilution) reactive with the N-terminal 4 Kd portion of lacZ. Sera incubations were performed for 2 hours at room temperature. Bound antibody was detected by addition of ^{125}I-labeled Protein A and subsequent exposure to film for variable times ranging from 16 hours

TABLE 2

Antigen	Human M. tb Sera	Anti-lacZ Sera
TbH-29	45 Kd	45 Kd
TbH 30	No reactivity	29 Kd
TbH-32	12 Kd	12 Kd
TbH-33	16 Kd	16 Kd

Positive reaction of the recombinant human *M. tuberculosis* antigens with both the human *M. tuberculosis* sera and anti-lacZ sera indicate that reactivity of the human *M. tuberculosis* sera is directed towards the fusion protein. Antigens reactive with the anti-lacZ sera but not with the human *M. tuberculosis* sera may be the result of the human *M. tuberculosis* sera recognizing conformational epitopes, or the antigen-antibody binding kinetics may be such that the 2 hour sera exposure in the immunoblot is not sufficient.

Studies were undertaken to determine whether the antigens TbH-9 and Tb38-1 represent cellular proteins or are secreted into *M. tuberculosis* culture media. In the first study, rabbit sera were raised against A) secretory proteins of *M. tuberculosis*, B) the known secretory recombinant *M. tuberculosis* antigen 85b, C) recombinant Tb38-1 and D) recombinant TbH-9, using protocols substantially as described in Example 3A. Total *M. tuberculosis* lysate, concentrated supernatant of *M. tuberculosis* cultures and the recombinant antigens 85b, TbH-9 and Tb38-1 were resolved on denaturing gels, immobilized on nitrocellulose membranes and duplicate blots were probed using the rabbit sera described above.

The results of this analysis using control sera raised [illegible]

residues and would therefore be expected to migrate with a mobility approximately 1 kD larger that the native protein. In Figure 2D, recombinant TbH-9 is lacking approximately 10 kD of the full-length 42 kD antigen, hence the significant difference in the size of the immunoreactive native TbH-9 antigen in the lysate lane (indicated by an arrow). These results demonstrate that Tb38-1 and TbH-9 are intracellular antigens and are not actively secreted by *M. tuberculosis*.

The finding that TbH-9 is an intracellular antigen was confirmed by determining the reactivity of TbH-9-specific human T cell clones to recombinant TbH-9, secretory *M. tuberculosis* proteins and PPD. A TbH-9-specific T cell clone (designated 131TbH-9) was generated from PBMC of a healthy PPD-positive donor. The proliferative response of 131TbH-9 to secretory proteins, recombinant TbH-9 and a control *M. tuberculosis* antigen, TbRa11, was determined by measuring uptake of tritiated thymidine, as described in Example 1. As shown in Figure 3A, the clone 131TbH-9 responds specifically to TbH-9, showing that TbH-9 is not a significant component of *M. tuberculosis* secretory proteins. Figure 3B shows the production of IFN-γ by a second TbH-9-specific T cell clone (designated PPD 800-10) prepared from PBMC from a healthy PPD-positive donor, following stimulation of the T cell clone with secretory proteins, PPD or recombinant TbH-9. These results further confirm that TbH-9 is not secreted by *M. tuberculosis*.

C. USE OF SERA FROM PATIENTS HAVING EXTRAPULMONARY TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). The resulting library was screened using pools of sera obtained from individuals with extrapulmonary tuberculosis [illegible]

153, respectively, with the 5' and 3' DNA sequences for XP32 being provided in SEQ ID NOS: 154 and 155, respectively. The predicted amino acid sequence for XP14 is provided in SEQ ID NO: 156. The reverse complement of XP14 was found to encode the amino acid sequence provided in SEQ ID NO: 157.

Comparison of the sequences for the remaining 14 clones (hereinafter referred to as XP1-XP6, XP17-XP19, XP22, XP25, XP27, XP30 and XP36) with those in the genebank as described above, revealed no homologies with the exception of the 3' ends of XP2 and XP6 which were found to bear some homology to known *M. tuberculosis* cosmids. The DNA sequences for XP27 and XP36 are shown in SEQ ID NOS: 158 and 159, respectively, with the 5' sequences for XP4, XP5, XP17 and XP30 being shown in SEQ ID NOS: 160-163, respectively, and the 5' and 3' sequences for XP2, XP3, XP6, XP18, XP19, XP22 and XP25 being shown in SEQ ID NOS: 164 and 165; 166 and 167, 168 and 169; 170 and 171; 172 and 173; 174 and 175; and 176 and 177, respectively. XP1 was found to overlap with the DNA sequences for TbH4, disclosed above. The full-length DNA sequence for TbH4-XP1 is provided in SEQ ID NO: 178. This DNA sequence was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 179. The reverse complement of TbH4-XP1 was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 180. The DNA sequence for XP36 was found to contain two open reading frames encoding the amino acid sequence shown in SEQ ID NOS: 181 and 182, with the reverse complement containing an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 183.

Recombinant XP1 protein was prepared as described above in Example 3B, with a metal ion affinity chromatography column being employed for purification. Recombinant XP1 was found to stimulate cell proliferation and IFN-γ production in T cells isolated from an *M. tuberculosis*-immune donors.

serological activity with a serum pool from *M. tuberculosis*-infected patients which showed little or no immunoreactivity with other antigens of the present invention. Rabbit anti-sera was generated against the most reactive fraction using the method described in Example 3A . The anti-sera was used to screen an *M. tuberculosis* Erdman strain genomic DNA expression library prepared as described above. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones determined.

Ten different clones were purified. Of these, one was found to be TbRa35, described above, and one was found to be the previously identified *M. tuberculosis* antigen, HSP60. Of the remaining eight clones, six (hereinafter referred to as RDIF2, RDIF5, RDIF8, RDIF10, RDIF11 and RDIF12) were found to bear some similarity to previously identified *M. tuberculosis* sequences. The determined DNA sequences for RDIF2, RDIF5, RDIF8, RDIF10 and RDIF11 are provided in SEQ ID NOS: 184-188, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NOS: 189-193, respectively. The 5' and 3' DNA sequences for RDIF12 are provided in SEQ ID NOS: 194 and 195, respectively. No significant homologies were found to the antigen RDIF-7. The determined DNA and predicted amino acid sequences for RDIF7 are provided in SEQ ID NOS: 196 and 197, respectively. One additional clone, referred to as RDIF6 was isolated, however, this was found to be identical to RDIF5.

Recombinant RDIF6, RDIF8, RDIF10 and RDIF11 were prepared as described above. These antigens were found to stimulate cell proliferation and IFN-γ production in T cells isolated from *M. tuberculosis*-immune donors.

EXAMPLE 4

PPD was prepared as published with some modification (Seibert, F. et al., Tuberculin purified protein derivative. Preparation and analyses of a large quantity for standard. The American Review of Tuberculosis 44:9-25, 1941). *M. tuberculosis* Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100°C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22 μ filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP-HPLC) using a C18 column (7.8 x 300 mM; Waters, Milford, MA) in a Biocad HPLC system (Perseptive Biosystems, Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% TFA in acetonitrile). The flow rate was 10 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80 μl/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from the

[illegible] Four cyanogen bromide fragments of DPPD were

EXAMPLE 5

SYNTHESIS OF SYNTHETIC POLYPEPTIDES

Polypeptides may be synthesized on a Millipore 9050 peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

This procedure was used to synthesize a TbM-1 peptide that contains one and a half repeats of a TbM-1 sequence. The TbM-1 peptide has the sequence GCGDRSGGNLDQIRLRRDRSGGNL (SEQ ID NO: 63).

EXAMPLE 6

USE OF REPRESENTATIVE ANTIGENS FOR SERODIAGNOSIS OF TUBERCULOSIS

This Example illustrates the diagnostic properties of several representative

[illegible] The plates were then removed and the wells were blocked for [illegible]

times with PBS/0.1% Tween 20™. 50 µL sera, diluted 1:100 in PBS/0.1% Tween 20™/0.1% BSA, was then added to each well and incubated for 30 minutes at room temperature. The plates were then washed again five times with PBS/0.1% Tween 20™.

The enzyme conjugate (horseradish peroxidase - Protein A, Zymed, San Francisco, CA) was then diluted 1:10,000 in PBS/0.1% Tween 20™/0.1% BSA, and 50 µL of the diluted conjugate was added to each well and incubated for 30 minutes at room temperature. Following incubation, the wells were washed five times with PBS/0.1% Tween 20™. 100 µL of tetramethylbenzidine peroxidase (TMB) substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD) was added, undiluted, and incubated for about 15 minutes. The reaction was stopped with the addition of 100 µL of 1 N H_2SO_4 to each well, and the plates were read at 450 nm.

Figure 4 shows the ELISA reactivity of two recombinant antigens isolated using method A in Example 3 (TbRa3 and TbRa9) with sera from *M. tuberculosis* positive and negative patients. The reactivity of these antigens is compared to that of bacterial lysate isolated from *M. tuberculosis* strain H37Ra (Difco, Detroit, MI). In both cases, the recombinant antigens differentiated positive from negative sera. Based on cut-off values obtained from receiver-operator curves, TbRa3 detected 56 out of 87 positive sera, and TbRa9 detected 111 out of 165 positive sera.

Figure 5 illustrates the ELISA reactivity of representative antigens isolated using method B of Example 3. The reactivity of the recombinant antigens TbH4, TbH12, Tb38-1 and the peptide TbM-1 (as described in Example 4) is compared to that of the 38 kD antigen described by Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989. Again, all of the polypeptides tested differentiated positive from negative sera. Based on cut-off values obtained from receiver-operator curves, TbH4 detected 67 out of 126 positive sera, TbH12 detected 50 out of 125 positive sera, 38-1 detected 61 out of 101 positive sera and the TbM-1

to the reactivity of *M. tuberculosis* lysate and the 38 kD antigen. The results are presented in Table 3, below:

TABLE 3

REACTIVITY OF ANTIGENS WITH SERA FROM *M. TUBERCULOSIS* PATIENTS

Patient	Acid Fast Sputum	ELISA Values					
		Lysate	38kD	TbRa9	TbH12	TbH4	TbRa3
Tb01B93I-2	++++	1.853	0.634	0.998	1.022	1.030	1.314
Tb01B93I-19	++++	2.657	2.322	0.608	0.837	1.857	2.335
Tb01B93I-8	+++	2.703	0.527	0.492	0.281	0.501	2.002
Tb01B93I-10	+++	1.665	1.301	0.685	0.216	0.448	0.458
Tb01B93I-11	+++	2.817	0.697	0.509	0.301	0.173	2.608
Tb01B93I-15	+++	1.28	0.283	0.808	0.218	1.537	0.811
Tb01B93I-16	+++	2.908	>3	0.899	0.441	0.593	1.080
Tb01B93I-25	+++	0.395	0.131	0.335	0.211	0.107	0.948
Tb01B93I-87	+++	2.653	2.432	2.282	0.977	1.221	0.857
Tb01B93I-89	++	1.912	2.370	2.436	0.876	0.520	0.952
Tb01B94I-108	+++	1.639	0.341	0.797	0.368	0.654	0.798
Tb01B94I-201	+++	1.721	0.419	0.661	0.137	0.064	0.692
Tb01B93I-88	++	1.939	1.269	2.519	1.381	0.214	0.530
Tb01B93I-92							

Patient	Acid Fast Sputum	ELISA Values					
		Lysate	38kD	TbRa9	TbH12	TbH4	TbRa3
Tb01B93I-9	+	2.649	0.278	0.210	0.140	0.181	1.586
Tb01B93I-14	+	>3	1.538	0.282	0.291	0.549	2.880
Tb01B93I-21	+	2.645	0.739	2.499	0.783	0.536	1.770
Tb01B93I-22	+	0.714	0.451	2.082	0.285	0.269	1.159
Tb01B93I-31	+	0.956	0.490	1.019	0.812	0.176	1.293
Tb01B93I-32	–	2.261	0.786	0.668	0.273	0.535	0.405
Tb01B93I-52	–	0.658	0.114	0.434	0.330	0.273	1.140
Tb01B93I-99	–	2.118	0.584	1.62	0.119	0.977	0.729
Tb01B94I-130	–	1.349	0.224	0.86	0.282	0.383	2.146
Tb01B94I-131	–	0.685	0.324	1.173	0.059	0.118	1.431
AT4-0070	Normal	0.072	0.043	0.092	0.071	0.040	0.039
AT4-0105	Normal	0.397	0.121	0.118	0.103	0.078	0.390
3/15/94-1	Normal	0.227	0.064	0.098	0.026	0.001	0.228
4/15/93-2	Normal	0.114	0.240	0.071	0.034	0.041	0.264
5/26/94-4	Normal	0.089	0.259	0.096	0.046	0.008	0.053
5/26/94-3	Normal	0.139	0.093	0.085	0.019	0.067	0.01

Based on cut-off values obtained from receiver-operator curves, TbRa3 detected 23 out of 27 positive sera, TbRa9 detected 22 out of 27, [illegible]

[illegible] detected [illegible] In addition, several of the recombinant antigens detected positive sera that were not detected by [illegible]

The reactivity of the recombinant antigen TbRa11 with sera from *M. tuberculosis* patients shown to be negative for the 38 kD antigen, as well as with sera from PPD positive and normal donors, was determined by ELISA as described above. The results are shown in Figure 6 which indicates that TbRa11, while being negative with sera from PPD positive and normal donors, detected sera that were negative with the 38 kD antigen. Of the thirteen 38 kD negative sera tested, nine were positive with TbRa11, indicating that this antigen may be reacting with a sub-group of 38 kD antigen negative sera. In contrast, in a group of 38 kD positive sera where TbRa11 was reactive, the mean OD 450 for TbRa11 was lower than that for the 38 kD antigen. The data indicate an inverse relationship between the presence of TbRa11 activity and 38 kD positivity.

The antigen TbRa2A was tested in an indirect ELISA using initially 50 µl of serum at 1:100 dilution for 30 minutes at room temperature followed by washing in PBS Tween and incubating for 30 minutes with biotinylated Protein A (Zymed, San Francisco, CA) at a 1:10,000 dilution. Following washing, 50 µl of streptavidin-horseradish peroxidase (Zymed) at 1:10,000 dilution was added and the mixture incubated for 30 minutes. After washing, the assay was developed with TMB substrate as described above. The reactivity of TbRa2A with sera from *M. tuberculosis* patients and normal donors in shown in Table 4. The mean value for reactivity of TbRa2A with sera from *M. tuberculosis* patients was 0.444 with a standard deviation of 0.309. The mean for reactivity with sera from normal donors was 0.109 with a standard deviation of 0.029. Testing of 38 kD negative sera (Figure 7) also indicated that the TbRa2A antigen was capable of detecting sera in this category.

TABLE 4

REACTIVITY OF TBRA2A WITH SERA FROM *M. TUBERCULOSIS* PATIENTS AND FROM NORMAL DONORS

[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

Tb91	TB	0.393
Tb92	TB	0.401
Tb93	TB	0.232
Tb94	TB	0.333
Tb95	TB	0.435
Tb96	TB	0.284
Tb97	TB	0.320
Tb99	TB	0.328
Tb100	TB	0.817
Tb101	TB	0.607
Tb102	TB	0.191
Tb103	TB	0.228
Tb107	TB	0.324
Tb109	TB	1.572
Tb112	TB	0.338
DL4-0176	Normal	0.036
AT4-0043	Normal	0.126
AT4-0044	Normal	0.130
AT4-0052	Normal	0.135
AT4-0053	Normal	0.133
AT4-0062	Normal	0.128
AT4-0070	Normal	0.088
AT4-0091	Normal	0.108
AT4-0100	Normal	0.106
AT4-0105	Normal	0.108
AT4-0109	Normal	0.105

The reactivity of the recombinant antigen (g) (SEQ ID NO: 60) with sera from *M. tuberculosis* patients and normal donors was determined by ELISA as described above. Figure 8 shows the results of the titration of antigen (g) with four *M. tuberculosis* positive sera that were all reactive with the 38 kD antigen and with four donor sera. All four positive sera were reactive with antigen (g).

The reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with

donors and with a pool of sera from *M. tuberculosis* patients. The mean OD 450 was demonstrated to be higher with sera from *M. tuberculosis* patients than from normal donors, with the mean OD 450 being significantly higher in the indirect ELISA than in the direct ELISA. Figure 11 is a titration curve for the reactivity of recombinant TbH-33 with sera from *M. tuberculosis* patients and from normal donors showing an increase in OD 450 with increasing concentration of antigen.

The reactivity of the recombinant antigens RDIF6, RDIF8 and RDIF10 (SEQ ID NOS: 184-187, respectively) with sera from *M. tuberculosis* patients and normal donors was determined by ELISA as described above. RDIF6 detected 6 out of 32 *M. tuberculosis* sera and 0 out of 15 normal sera; RDIF8 detected 14 out of 32 *M. tuberculosis* sera and 0 out of 15 normal sera; and RDIF10 detected 4 out of 27 *M. tuberculosis* sera and 1 out of 15 normal sera. In addition, RDIF10 was found to detect 0 out of 5 sera from PPD-positive donors.

EXAMPLE 7

PREPARATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* FUSION PROTEINS

A fusion protein containing TbRa3, the 38 kD antigen and Tb38-1 was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR in order to facilitate their fusion and the subsequent expression of the fusion protein TbRa3-38 kD-Tb38-1. TbRa3, 38 kD and Tb38-1 DNA was used to perform PCR using the primers PDM-64 and PDM-65 (SEQ ID NO: 141 and 142), PDM-57 and PDM-58 (SEQ ID NO: 143 and 144), and PDM-69 and PDM-60 (SEQ ID NO: 145-146), respectively. In each case, the DNA amplification was performed using 10 µl 10X Pfu buffer, 2 µl 10 mM dNTPs, 2 µl each of the PCR primers [illegible]

68°C for 15 sec and 72°C for 3 min, and finally by 72°C for 4 min. For Tb38-1 denaturation at 94°C for 2 min was followed by 10 cycles of 96°C for 15 sec, 68°C for 15 sec and 72°C for 1.5 min, 30 cycles of 96°C for 15 sec, 64°C for 15 sec and 72°C for 1.5, and finally by 72°C for 4 min.

The TbRa3 PCR fragment was digested with NdeI and EcoRI and cloned directly into pT7^L2 II. 1 vector using NdeI and EcoRI sites. The 38 kD PCR fragment was digested with Sse8387I, treated with T4 DNA polymerase to make blunt ends and then digested with EcoRI for direct cloning into the pT7^L2Ra3-1 vector which was digested with StuI and EcoRI. The 38-1 PCR fragment was digested with Eco47III and EcoRI and directly subcloned into pT7^L2Ra3/38kD-17 digested with the same enzymes. The whole fusion was then transferred to pET28b using NdeI and EcoRI sites. The fusion construct was confirmed by DNA sequencing.

The expression construct was transformed to BLR pLys S *E. coli* (Novagen, Madison, WI) and grown overnight in LB broth with kanamycin (30 μg/ml) and chloramphenicol (34 μg/ml). This culture (12 ml) was used to inoculate 500 ml 2XYT with the same antibiotics and the culture was induced with IPTG at an OD560 of 0.44 to a final concentration of 1.2 mM. Four hours post-induction, the bacteria were harvested and sonicated in 20 mM Tris (8.0), 100 mM NaCl, 0.1% DOC, 20 μg/ml Leupeptin, 20 mM PMSF followed by centrifugation at 26,000 X g. The resulting pellet was resuspended in 8 M urea, 20 mM Tris (8.0), 100 mM NaCl and bound to Pro-bond nickel resin (Invitrogen, Carlsbad, CA). The column was washed several times with the above buffer then eluted with an imidazole gradient (50 mM, 100 mM, 500 mM imidazole was added to 8 M urea, 20 mM Tris (8.0), 100 mM NaCl). The eluates containing the protein of interest were then dialzyed against 10 mM Tris (8.0)

The DNA and amino acid sequences for the resulting fusion protein

procedure to that described above. The DNA sequence for the TbH9-Tb38-1 fusion protein is provided in SEQ ID NO: 151.

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and DPEP was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR and cloned into vectors essentially as described above, with the primers PDM-69 (SEQ ID NO:145 and PDM-83 (SEQ ID NO: 200) being used for amplification of the Tb38-1A fragment. Tb38-1A differs from Tb38-1 by a DraI site at the 3' end of the coding region that keeps the final amino acid intact while creating a blunt restriction site that is in frame. The TbRa3/38kD/Tb38-1A fusion was then transferred to pET28b using NdeI and EcoR1 sites.

DPEP DNA was used to perform PCR using the primers PDM-84 and PDM-85 (SEQ ID NO: 201 and 202, respectively) and 1 µl DNA at 50 ng/µl. Denaturation at 94 °C was performed for 2 min, followed by 10 cycles of 96 °C for 15 sec, 68 °C for 15 sec and 72 °C for 1.5 min; 30 cycles of 96 °C for 15 sec, 64 °C for 15 sec and 72 °C for 1.5 min; and finally by 72 °C for 4 min. The DPEP PCR fragment was digested with EcoRI and Eco72I and clones directly into the pET28Ra3/38kD/38-1A construct which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-2) are provided in SEQ ID NO: 203 and 204, respectively.

EXAMPLE 8

USE OF *M. TUBERCULOSIS* FUSION PROTEINS FOR SERODIAGNOSIS OF TUBERCULOSIS

The effectiveness of the fusion protein TbRa3-38 kD-Tb38-1, prepared as described above, in the serodiagnosis of tuberculosis infection was examined by ELISA.

The ELISA protocol was as described above in Example 6, with the fusion

the three antigens individually or in combination. Such a panel enabled the dissection of the serological reactivity of the fusion protein to determine if all three epitopes functioned with the fusion protein. As shown in Table 5, all four sera that reacted with TbRa3 only were detectable with the fusion protein. Three sera that reacted only with Tb38-1 were also detectable, as were two sear that reacted with 38 kD alone. The remaining 15 sera were all positive with the fusion protein based on a cut-off in the assay of mean negatives +3 standard deviations. This data demonstrates the functional activity of all three epitopes in the fusion protein.

TABLE 5

REACTIVITY OF TRI-PEPTIDE FUSION PROTEIN WITH SERA FROM *M. TUBERCULOSIS* PATIENTS

Serum ID	Status	ELISA and/or Western Blot Reactivity with Individual proteins			Fusion recombinant OD 450	Fusion Recombinant Status
		38kd	Tb38-1	TbRa3		
01B93I-40	TB	-	-	+	0.413	+
01B93I-41	TB	-	+	+	0.392	+
01B93I-29	TB	+	-	+	2.217	+
01B93I-109	TB	+	+	+	0.522	+
01B93I-132	TB	+	+	+	0.937	+
5004	TB	+	+	+	1.098	+
15004	TB	+	+	+	2.077	+
39004	TB	+	+	+	1.675	+
68004	TB	+	+	+	2.388	+
99004	TB	-	+	+	0.607	-
107004	TB	-	+	+	0.667	-
92004	TB	+	+	+	1.070	+
97004	TB	+	-	+	1.152	+
118004	TB	+	-	+	2.694	+
173004	TB	+	+	+	3.258	+
175004	TB	+	-	+	2.514	+
274004	TB					

289004	TB	-	-	+	0.848	+
308004	TB	-	+	-	3.338	+
314004	TB	-	+	-	1.362	+
317004	TB	+	-	-	0.763	+
312004	TB	-	-	+	1.079	+
D176	PPD	-	-	-	0.145	-
D162	PPD	-	-	-	0.073	-
D161	PPD	-	-	-	0.097	-
D27	PPD	-	-	-	0.082	-
A6-124	NORMAL	-		-	0.053	-
A6-125	NORMAL	-	-	-	0.087	-
A6-126	NORMAL	-	-	-	0.346	+
A6-127	NORMAL	-	-	-	0.064	-
A6-128	NORMAL	-	-	-	0.034	-
A6-129	NORMAL	-	-	-	0.037	-
A6-130	NORMAL	-	-	-	0.057	-
A6-131	NORMAL	-	-	-	0.054	-
A6-132	NORMAL	-	-		0.022	-
A6-133	NORMAL	-	-		0.147	-
A6-134	NORMAL	-	-	-	0.101	-
A6-135	NORMAL	-	-		0.066	-
A6-136	NORMAL	-	-		0.054	-
A6-137	NORMAL	-	-	-	0.065	-
A6-138	NORMAL	-	-	-	0.041	-
A6-139	NORMAL	-			0.103	
A6-140	NORMAL	-	-	-	0.212	-
A6-141	NORMAL		-	-	0.056	-
A6-142	NORMAL	-	-	-	0.051	-

The reactivity of the fusion protein Tbl -2 with sera from *M. tuberculosis*-infected patients was examined by ELISA using the protocol described above. The results of these studies (Table 6) demonstrate that

TABLE 6

REACTIVITY OF TbF-2 FUSION PROTEIN WITH TB AND NORMAL SERA

Serum ID	Status	TbF OD450	Status	TbF-2 OD450	Status	ELISA Reactivity			
						38 kD	TbRa3	Tb38-1	DPEP
B931-40	TB	0.57	+	0.321	+	-	+	-	+
B931-41	TB	0.601	+	0.396	+	+	+	+	-
B931-109	TB	0.494	+	0.404	+	+	+	±	-
B931-132	TB	1.502	+	1.292	+	+	+	+	±
5004	TB	1.806	+	1.666	+	±	+	+	-
15004	TB	2.862	+	2.468	+	+	+	+	-
39004	TB	2.443	+	1.722	+	+	+	+	-
68004	TB	2.871	+	2.575	+	+	+	+	-
99004	TB	0.691	+	0.971	+	-	+	+	-
107004	TB	0.875	+	0.732	+	-	±	+	-
92004	TB	1.632	+	1.394	+	+	±	±	-
97004	TB	1.491	+	1.979	+	+	±	-	+
118004	TB	3.182	+	3.045	+	+	±	-	-
173004	TB	3.644	+	3.578	+	+	+	+	-
175004	TB	3.332	+	2.916	+	+	+	-	-
274004	TB	3.696	+	3.716	+	-	+	-	+
276004	TB	3.243	+	2.56	+	-	-	+	-
282004	TB	1.249	+	1.234	+	+	-	-	-
289004	TB	1.373	+	1.17	+	-	+	-	-
308004	TB	3.708	+	3.355	+	-	-	+	-
314004	TB	1.663	+	1.399	+	-	-	+	-
317004	TB	1.163	+	0.92	+	+	-	-	-
312004	TB	1.709	+	1.453	+	-	+	-	-
380004	TB	0.238	-	0.461	+	-	+	-	+
451004	TB	0.18	-	0.2	-	-	-	-	+
478004	TB	0.188	-	0.469	+	-	-	-	+
410004	TB	0.384	+	2.392	+	+	-	-	+
411004	TB	0.306	+	0.874	+	-	+	-	+
421004	TB	0.357	+	1.456	+	-	+	-	+
528004	TB	0.047	-	0.196	+	-	-	-	+
A6-87	Normal	0.094		0.063		-	-	-	-
A6-88	Normal	0.214		0.19	-	-	-	-	-
A6-89	Normal	0.248		0.125	-	-	-	-	-
A6-90	Normal	0.179	-	0.206	-	-	-	-	-
A6-91	Normal	0.135	-	0.151			-	-	
A6-92	Normal	0.064		0.097		-	-	-	
A6-93	Normal	0.072	-	0.098			-	-	
A6-94	Normal	0.072		0.064					
A6-95	Normal	0.125		0.159					

One of skill in the art will appreciate that the order of the individual antigens within the fusion protein may be changed and that comparable activity would be expected provided each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

5

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.

(ii) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

(iii) NUMBER OF SEQUENCES: 209

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SEED and BERRY LLP
(B) STREET: 6300 Columbia Center, 701 Fifth Avenue
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: 98104-7092

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 01-OCT-1997
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/027,824
(B) FILING DATE: 04-OCT-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Lee, Lydia T.
(B) REGISTRATION NUMBER: 31,778
(C) REFERENCE/DOCKET NUMBER: 210.1.1

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (206) 622-4900
(B) TELEFAX: (206) 682-6031

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGAGGCACCG GTAGTTTGAA CCAAACGCAC AATCGACGGG CAAACGAACG GAAGAACACA 60

ACCATGAAGA TGGTGAAATC CATCGCCGCA GGTCTGACCG CCGCGGCTGC AATCGGCGCC 120

GCTGCGGCCG GTGTGACTTC CATCATGGCT GCCGGCCCGG TCGTATACCA GATGCAGCCG 180

GTCGTCTTCG GCGCGCCACT GCCGTTGGAC CCGGCATCCG CCTCTGACGT CCCGACCGCC 240

GCCCAGTTGA CCAGCTGCCT CAACAGCCTC GCCGATCCCA ACGTGTCGTT TGCGAACAAG 300

GGCAGTCTGG TCGAGGGCGG CATCGGGGGC ACCGAGGCGC GCATCGCCGA CCACAAGCTG 360

AAGAAGCCCG CCGACGACGG GGATCTGCCG CTGTCGTTCA GCGTGACGAA CATCCAGCCG 420

GCGGCCGCCG GTTCGGCCAC CGTCGACGTT TCCGTCTCGG GTCCGAAGCT CTCGTCGCCG 480

GTCACGCAGA ACGTCACGTT CGTGAATCAA GGCGGCTGGA TGCTGTCACG CGCATCGGGG 540

ATGGAGTTGC TGCAGGCCGC AGGGNAACTG ATTGNCGGGC CGGNTTCAGC CCGCTGTTCA 600

GCTACGCCGC CCGCCTGGTG ACGCGTCCAT GTCGAACACT CGCGCGTGTA GCACCGTGCG 660

GTNTGCGCAG CGNCGCACGC ACCGCCCGGT GTAAGCCGTC GTCGAGATAG GTGGTGNCTC 720

GNTACCAGNG ANCACCGCCG NNTCGNCNTT TCTCGNTGTT GNATGA 766

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 792 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

TTTCTCGACG ACGTGACCGT GAGCCGTCGC CATGCTGAAT TCCCGTTGGA AAACAACGAA 300

TTCAATGTCG TCGATGTCGG GAGTCTCAAC GGCACCTACC TCAACCGCGA GCCCGTGGAT 360

TCGGCGGTGC TGGCGAACGG CGACGAGGTC CAGATCGGCA AGCTCCGGTT GGTGTTCTTG 420

ACCGGACCCA AGCAAGGCGA GGATGACGGG AGTACCGGGG GCCCGTGAGC GCACCCGATA 480

GCCCCGCGCT GGCCGGGATG TCGATCGGGG CGGTCCTCCG ACCTGCTACG ACCGGATTTT 540

CCCTGATGTC CACCATCTCC AAGATTCGAT TCTTGGGAGG CTTGAGGGTC NGGGTGACCC 600

GCCCGCGGGC CTCATTCNGG GGTNTCGGCN GGTTTCACCC CNTACCNACT GCCNCCTGGN 660

TTGCNAATTC NTTCTTCNGT GCCCNNAAAG GGACCNTTAN CTTGCCGCIN GAAANGGTNA 720

TCCNGGGCCC NTCCTNGAAN CCCCNTCCCC CT 752

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 813 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CATATGCATC ACCATCACCA TCATACTTCT AAAGGCTAAC CCTCTCGGGG GCGTCGAGCA 60

CCAGCCGACA CCGGGCCGGA TCGATCTGCT ATTTGACTC GCTCAGGTA TGGTCGTCAG 120

CAGTCCGATG CGCTATGTTT GTCGTCGACT CAGATATCGC GCAATCCAA TCTCCCGCCT 180

[illegible] 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

[illegible] 540

GAGCAACGCA GACCGGGACA ACWGGTATCG ATAGCGGCCN AATGCCGGCT TGGAACCCNG 780

TGAAATTATC ACAACTTCGC AGTCACNAAA NAA 813

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGCTATGAAC ACGGCCGCGT CCGATAACTT CGACGTGTCC CAGGGTGGGC AGGGATTCGC 60

CATTCCGATC GGGCAGGCGA TGGCGATCGC GGGCCAGATC CGATCGCGTG GGCGCTCACC 120

CACCGTTCAT ATCGGGCCTA CCGCCTTCCT CGGCTTGGGT GTTGTCGACA ACAACGGCAA 180

CGGCGCACGA GTCCAACCGG TGGTCGGGAC CGGTCGGGCG GCAAGTCTCG GCATCTCCAC 240

CGGCGACGTG ATCACCGCGG TCGACGGCGC TCCGATCAAC TCGGCCACCG CGATGGCGGA 300

CGCGCTTAAC CGGCATCAAT CGGCCGACGT CATTCTGGTG AACTGGCAAA CGAAGTCGGG 360

CGGCAGCCTT ACAGCGAACT TGACATTCGT CGAAGGACCA CCGAGTTTAT TTCGTGGTGG 420

ATACCATTTT CGGCCTTTTT AATTGGA 447

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

CCGGCGANGG NGAGCGCCGG AATGGCGCGA GTGAGGAGGT GGNCAGTCAT GCCCAGNGTG 240

ATCCAATCAA CCTGNATTCG GNCTGNGGGN CCATTTGACA ATCGAGGTAG TGAGCGCAAA 300

TGAATGATGG AAAACGGGNG GNGACGTCCG NTGTTCTGGT GGTGNTAGGT GNCTGNCTGG 360

NGTNGNGGNT ATCAGGATGT TCTTCGNCGA AANCTGATGN CGAGGAACAG GGTGTNCCCG 420

NNANNCCNAN GGNGTCCNAN CCCNNNNTCC TCGNCGANAT CANANAGNCG NTTGATGNGA 480

NAAAAGGGTG GANCAGNNNN AANTNGNGGN CCHAANAANC NNNANNCNNG NNAGNTNGNT 540

NNNTNTTNNC ANNNNNNNTG NNGNNGNNCN NNNCAANCNN NTNNNNGNAA NNGGNTTNTT 600

NAAT 604

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 633 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTGCANGTCG AACCACTCGA CTAAACGGAA CAAAAGCTGG AGTTCACTCG TGGTGCGGGC 60

CCTCTTAGAA CTACTCKATC YTTCKCGTTC TAAGATTCCC GYACGAGCAT TAKGACAGTC 120

TAANCTCCTC GTTACGTTGA TTGAATAAGC AAGAATTCCC TCTGATCGCA AATTAACGAA 180

CGAGTCAGAA CCTCAAATCT CAACTCGTTC CATTTCGTCA AAGGGTCTTC GTCCTCGTTA 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 633

(A) LENGTH: 1362 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGACGACGAC GGCGCCGCAG AGCGGGCGCG AACGGCGATC GACGCGGCCC TGGCCAGAGT 60

CGGCACCACC CAGGAGGGAG TCGAATCATG AAATTTGTCA ACGATATTGA GCCCGTCGCG 120

CCCCGCCGAG CCGGCGGCGC GGTCGCCGAG GTCTATGCCG ACCCCCGCCT CGAGTTCGGC 180

CGGCTGCTCG AGCCCGTCGC CATGCCGTCC CCGGACGAGG GACTGCTCAC CGCCGGCTGG 240

GCGACGTTGC GCGAGACACT GTTGGCGGGC CAGGTGCCGC GTGGCCGCAA GGAAGCGGTC 300

GCCGCCGCCG TCGCGCGCCAG CCTGCGCTGC CCCTGGTGCG TCGACGCACA CACCACCATG 360

CTGTACGCGG CAGGCCAAAC CGACACCGCC GCGGCGATCT TGGCCGGCAC AGCACCTGCC 420

GCCGGTGACC CGAACGCGCC GTATGTGGCG TGGGCGGCAG GAACCGGGAC ACCGGCGGGA 480

GCGCCGGCAC CGTTCGGCCC GGATGTCGCC GCCGAATACC TGGGCACCGG GGTGCAATTT 540

CACTTCATCG CACCCCTGGT CCTGGTGCTC CTGGACGAAA CCTTCCTGCC GGGGGGCCCG 600

CGCGCGAAC AGCTCATCGA CCTGCTCGCC GGACTGGTGT TCTTCCCCAA CGTCGCCCTG 660

GAGCATCCGC CGGCCGGTTC CACCGCGAG CTGGACGCCC GAACGCTCC CGACATCTG 720

GCATCGGCAA CACCGTCCGA CTCGATAGCA ACCTCGTTCG CCTTCCTCAA CGACCACCTG 780

GACACCGCGC CGCACCTGCC GCTACCGACT TCTCAGCTCC TCAACGCCGC CTTCGCGTCG 840

GGACCGGACG GTCACCGGGG GTCACCCTGC GCGGCCAAGG AA 1362

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCGACGACCC CGATATGCCG CGCACCGTAG CGAAAGCCGT CGCCGACGCA GTCGGGCGCG 60

GTATCGCTCC CGTTGAGGAG ATTCAGGACT GCGTGGAGGC CTGGCTGGCG GAAGCCGGTC 120

TGGATGACGT GGCCCGTGTT TACATCATCT ACCGGCAGCG GCGCGCCGAG CTGCGGACGG 180

CTAAGGCCTT GCTCGGCGTG CGGGACGAGT TAAAGCTGAG CTTGGCGGCC GTGACGGTAC 240

TGCGCGAGCG CTATCTGCTG CACGACGAGC ACGGCCGGCC GGCCGAGTCG ACCGGCGAGC 300

TGATGGACCG ATCGGCGCGC TGTGTCGCGG CGGCCGAGGA CCAGTATGAG CCGGGCTCGT 360

CGAGGCGGTG GGCCGAGCGG TTCGCCAGGC TATTACGCAA CCTGGAATTC CTGCCGAATT 420

CGCCCACGTT GATGAACTTT GTTACTGACT TTGACTCTGT CGTCGGCTCT TTTGTTCTGC 480

TGATTGAGGA TTTCTTCAAA TCGACTTTT CGACTTCGGG ACAGGCCGCG GAGCTGCAGC 540

GGGCTGGAGG CGGTACTGA TATCGTTTA TTACTTCGC ATCCGCCGGG TATCGGGTGG 600

CCTCCACGGC CGGCACGGCG AGCGAACTGG TGTTTCTT ATCGCTTAT GAAATGCCCC 660

[illegible] 720

[illegible] 780

[illegible] 840

[illegible] 900

[illegible] 960

[illegible] 1020

CGGCCCGCGC CACCCGCAAG ATCGGGCTCG GAGTCATGGG TTTGGCGGAA CTGCTTGCCG 1260

CACTGGGTAT TCCGTACGAC AGTGAAGAAG CCGTGCGGTT AGCCACCCGG CTCATGCGTC 1320

GCATACAGCA GGCGGCGCAC ACGGCATCGC GGAGGCTGGC CGAAGAGCGG GGCGCATTCC 1380

CGGCGTTCAC CGATAGCCGG TTCGCGCGGT CGGGCCCGAG GCGCAACGCA CAGGTCACCT 1440

CCGTCGCTCC GACGGGCA 1458

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 862 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACGGTGTAAT CGTGCTGGAT CTGGAACCGT GTGGCAGCGT ACGTACCGAC ATGTACTGGC 60

GGCGCAGGGG GCTGGCCTTG GGCATCGCGG TCGTCGTAGT CGGGATCGCG CTGGCCATCG 120

TCATCCCCTT CGTTGACAGC AGCGCCGTCG CGAAACCGGT CAGCGCCGAC AAGCCGGCCT 180

CGCGTCAGAG CATTCGTGTC TCGCCGGTAC CCAACGTAGT CCAGCCGGCC GGGCAAACCG 240

AACGTAACGC CGTCGCGTCC CTGCCCAAGT CTAAAACCGT CAGCACGCCT ACGCTACGCA 300

CTGCGGTGCA CGCGCCGTCG GTCTTAAAGA AAGCGACTA TTCGTCGAAT TGACCTTCG 360

CCGTTAAAGG TTTAAAAAC CCCATTGAT GTGCTTCGT CGATAGCGTC AACTTCACCA 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 660

[illegible] 720

[illegible] 780

[illegible] 840

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTGATCAGCA CCGGCAAGGC GTCACATGCC TCCCTGGGTG TGCAGGTGAC CAATGACAAA 60

GACACCTCGG GCGCGAAGAT CGTCGAAGTA GTGGCCGGTG CTGCTGCCGC GAACGCTGGA 120

GTCCTGAAGG GCGTCGTTGC TAACTAAGTT GACGACCGCC CGATCAACAG TGCGGACGCG 180

TTGGTTGCGC CCGTGCGGTC CAAAGCGCCC GGCGCCACGG TGCCGCTAAC CTTTCAGGAT 240

CCGTCGGGAG GTAGCGGCAC AGTGCAAGTC ACCCTCGGCA AGGCGGAGCA GTGATGAAGG 300

TCGCCGCGCA GTGTTCAAAG CTCGGATATA CGGTGGCACC CATGGAACAG CGTGCGGAGT 360

TGGTGGTTTG GCCGGGACTT GTCTTCGTCG TTGACGATCG CACGGCGCAC CGCGATGAAG 420

ACCACAGCTT CCCCTTTTTC ACCGAGCTTC TATCGAAGCC CGGGTTTCTT GTCGACGGCG 480

TGGTGGCGGT GTCCTTCGAA GACTTCGACA TCTGAAAATG CATGAACACA GCGGTGATCG 540

GCGGGGTGGA CCTGGTGGTG TCGCTCGCCA ACCCGGACCT GACGGTTTCC GATGTCACCG 600

CGAAGCCAC CTGGACATT CT 622

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

GCCTACGTGC GATCGTGCCC GGGCTACACG TTGGACTACA ACGCCAACGG GTCCGGTGCC 240

GGGGTGACCC AGTTTCTCAA CAACGAAACC GATTTCGCCG GCTCGGATGT CCCGTTGAAT 300

CCGTCGACCG GTCAACCTGA CCGGTCGGCG GAGCGGTGCG GTTCCCCGGC ATGGGACCTG 360

CCGACGGTGT TCGGCCCGAT CGCGATCACC TACAATATCA AGGGCGTGAG CACGCTGAAT 420

CTTGACGGAC CCACTACCGC CAAGATTTTC AACGGCACCA TCACCGTGTG GAATGATCCA 480

CAGATCCAAG CCCTCAACTC CGGCACCGAC CTGCCGCCAA CACCGATTAG CGTTATCTTC 540

CGCAGCGACA AGTCCGGTAC GTCGGACAAC TTCCAGAAAT ACCTCGACGG TGTATCCAAC 600

GGGGCGTGGG ACAAAGGCGC CAGCGAAACG TTCAGCGGGG GCGTCGGCGT CGGCGCCAGC 660

GGGAACAACG GAACGTCCGC CCTACTGCAG ACGACCGACG GTTCGATCAC CTACAAGGAG 720

TCGTCGTTTG CGGTGGGTAA GCAGTTGAAC ATGGCCCAGA TCATCACGTC GGCGGGTCCG 780

GATCCAGTGG CGATCACCAC CGAGTCGGTC CGTAAGACAA TCGCCGGGGC CAAGATCATT 840

GGACAAGGCA ACGACCTGGT ATTGGACACG TCGTCGTTCT ACAGACCCAC CCAGCCTGGT 900

TCTTACCCGA TCGTGCTGGC GACCTATGAG ATCGTCTGCT CGAAATACCC GGATGCGACG 960

ACCGGTACTG CGGTAAGGGC GTTTATGCAA GCCGCCATTG GTCCAGGCCA AGAAGGCCTG 1020

GACCAATACC ACTCCATTCC GTTGCGCAAA TCGTTTCAAG CAAAATTCCA GCCGCGCTTC 1080

AATGCTATTT CTTGACCTAG TGAAGGAAT TCGACGCTGA CGACGCCTT TCTGGAGGTA 1140

GCGTCCTAAT TTCGCCCTCA TAGCTATTC [illegible] 1200

(2) INFORMATION FOR SEQ ID NO:7 :

[illegible]

GGTTGCTCCA AGCGGTGGCC GCCGACGGCC GCATCCACAC CACGTTCAAC CAGACGATCG 240

CCGCGACCGG CCGGCTCTCC TCGACCGAAC CCAACCTGCA GAACATCCCG ATCCGCACCG 300

ACGCGGGCCG GCGGATCCGG GACGCGTTCG TGGTCGGGGA CGGTTACGCC GAGTTGATGA 360

CGGCCGACTA CAGCCAGATC GAGATGCCGA TCATGGGGCA CCTGTCCGGG GACGAGGGCC 420

TCATCGAGGC GTTCAACACC GGCGAGGACC TGTATTCGTT CGTCGCGTCC CGGGTGTTCG 480

GTCGTCGCTAT CGACGAGGTC ACCGGCGAGT TGCGGCGCCG GGTCAAGGCG ATGTCGTACG 540

GGCTCGTTTA CGGGTTGAGC GGCTACGGCC TGTCGCAGCA GTTGAAAATC TCCACCGAGG 600

AAGCGAACGA GCAGATGGAC GCGTATTTCG CCCGATTCGG CGGCGTCGCC GACTACCTGG 660

ACGCCGTAGT CGAGCGGGCC CGCAAGGACG GCTACACCTC GACGGTGCTG GGCCGTCGCC 720

GCTACCTGCC CGACCTGGAC AGCAGCGACC GTCAAGTGCG GGAGGCCGCC GAGTGGGCGA 780

CGCTGAACGC GCCCATCCAG GGCAGCGCGG CCGACATCAT CAAGGTGGCC ATGATCCAGG 840

TCGACAAGGC GCTCAACGAG GCACAGCTGG CGTCGCGCAT GCTGCTGCAG GTCCACGACG 900

AGCTGGTCTT CGAAATCGCC CCCGGTCAAG TCGACCGGCT CGAGGCGCTG GTGCGCGACA 960

AGATGGAGGG CGCTTACCCG CTGGACCTCT CGCTGCAGGT GTCGGTGGGC TACGGCCGCA 1020

GCTGGGACGC GCACGCCCAC TCAGTGCCGA CGTTGCATCT CGGGCGCAAA TTCGGCGATT 1080

TTCGCGACCT AAGTTACGCC TCGCGTCAAC TGGACTCAAC TTTTTCCAGC GTGTACCCGT 1140

CGATACGCCT CGTCA 1155

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) STRAINDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

GGCCGTCGGA TGCCGATTCC GCACGTTTCC GGTGCGACGG CTGGCCCTTC GAGCACGGAC 240

ATCGAGAACT CTCGGGGTTC GGCGAACGTT ATCTCAGTGG AATCTCAGTC CACGCGCGCA 300

ACCTAGTTGT GCAGTTACTG TTGAAAGCCA CACCCATGCC AGTCCACGCA TGGCCAAGTT 360

GGCCCGAGTA GTGGGCCTAG TACAGGAAGA GCAACCTAGC GACATGACGA ATCACCCACG 420

GTATTCGCCA CCGCCGCAGC AGCCGGGAAC CCCAGGTTAT GCTCAGGGGC AGCAGCAAAC 480

GTACAGCCAG CAGTTCGACT GGCGTTACCC ACCGTCCCCG GCCCCGCAGC CAACCCAGTA 540

CGGTCAACCC TACGAGGCGT TGGGTGGTAC CCGGCCGGGT CTGATACTTG GCGTGATTCC 600

GACGTATGAG CCCCGTCCTG GGATGGTTCG CCAACGACCT CGTGCAGTCA TGTTGGCCAT 660

CGGCGGCGTG ACGATAGCGG TGGTGTCGGC CGGCATCGGC GGCGGCGCGG CATCGCTGGT 720

CGGGTTCAAC CGGCCAGCGG CGGTTACCAG CGGCGGCCCA GTGGCTGTTA GCGGCGGCGC 780

AAGCATCCGC GCAGCAAACA TGCCGCCGGG GTCGGTCGAA CAGGTGGCGG CCAAGGTGGT 840

GCGCAGTGTC GTCATGTTGG AAACCGATCT GGGCGCGCAG TCGGAGGACA GCTCGGGCAT 900

CATTCTGTCT GCCCAGGGGC TGATCTTGAC CAACAACCAC GTCATCGCGG CGGCCGCCAA 960

GCGTCCGCTG GGCAGTCCGC CGCCGAAAAC GACGGTAACC TTCTCTGACG GGCCGACCGC 1020

ACCCTTCATG GTGGTGGGGC CTGACCCCAC CAGTGATATC GCCGTCGTCC GTGTTCAGGG 1080

CGTCTCCGAG CTCAACCAGA TCTCCCTGGG TTCGTCCTCG GACCTGAGCA TCGGTGAGCC 1140

GGTGTTGGCG ATCGGCTCGC CGCTCGGTTT GGAGGGCACT GTCACCAGCG GCATCGTCAG 1200

CGCTCTTAAC AGTGTG CAACGACTTC CAACAGTCAC AACAGGACAA CTCCTCGGCA 1260

CGCCATTCAG ACCCACCGCG CAATCAACCC CGGTAACTCC GGAAGCCCTC TGTTTGAACAT 1320

GA [illegible] 1380

[illegible] 1440

[illegible] 1500

[illegible] 1560

[illegible] 1620

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1058 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCCACCGGG CTGGCGGCCG CTCTAGAACT AGTGGATCCC CCGGGCTGCA GGAATTCGGC 60

ACGAGGATCC CACGTCGCAG GTTGTCGAAC CCGCCGCCGC GAAGTATCG GTCCATCCCT 120

AGCTCGGCGA CGGCGAGCGC CGGAATGGCG CGAGTGAGGA GGCGGGCAAT TTGGCGGGGC 180

CCGGCGACGG CGAGGGCGGC AATGGCGCGA GTGAGGAGGC GGGCAGTCAT GCCCAGCGTG 240

ATCCAATCAA CCTGCATTCC GCCTGCGGGC CCATTTGACA ATCGAGGTAC TGAGCGCAAA 300

TGAATGATGG AAAACGGGCG GTGACGTCCG CTGTTCTGGT GGTGCTAGGT GCCTGCCTGG 360

CGTTGTGGCT ATCAGGATGT TCTTCGCCGA AACCTGATGC CGAGGAACAG GGTGTTCCCG 420

TGAGCCCGAC GGCGTCCGAC CGCGCGCTTG TGGCCGAGAT CAGGCAGTCG CTTGATGCGA 480

CAAAAGGTTT GACCAGCGTG CACGTAGGCG TCGAACAACT CGGGAAAGTC GACAGCTTGC 540

TGGGTATTAC CAGTGCCGAT GTGGACGTGC CGGGCAATCC GTCGCGCGCA AAGCGCGTAT 600

CGATTATAAA CGACGAGTAT GATTTTGATC TTCCAATTAA AAGAACAACC ATTATGGTGA 660

AACTGTTCGA CGACTGGAGT AATCTGGGTA TGATTATTAA ATCTAAACAT TACGCGCTGT 720

TCGATCCTGT TTTTGGTCTG AAGGATGTGC TGATTGGTTT TAGGAATTTT AAACGTAAAT 780

[illegible] 840

[illegible] 900

[illegible] 960

[illegible] 1020

[illegible] 1058

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCGGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAACAGGC 60

GGCGGCGGAG GCGGTCCAGC GGGCGCGGGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT 120

CATTGAGCAG GACATGGCCG TGGACAGCGT CGGCAAGATC ACCTACCGCA TCAAGCTCGA 180

AGTGTCGTTC AAGATGAGGC CGGCGCAACG GCGCTAGCAC GGGCCGGCGA GCAAGACGCA 240

AAATCGCACG GTTTGCGGTT GATTCGTGCG ATTTTGTGTC TGCTCGCCGA CGGCTAGCAG 300

GTGCGGTGCA GGTJCGGGTG CTGTAGTATC CAGGCGTGCA TCGCGATTCC GGCGGCCACG 360

CCGGAGTTAA TGCTTCGCGT CGACCCGAAC TGGGCGATCC GCCGGNGAGC TGATCGATGA 420

CCGTGGCCAG CCCGTCGATG CCCGAGTTGT CCGAGGAAAC GTCCTGCCAG GCCGGTAGGA 480

AGAGTCGGTA GGCGGCGGTG CTGACCGGCT GTGCGTGCGC GCTCAGTGCG GCCAGCGAGC 540

GG 542

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 814 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

GTTTGCCGCC AATATTCGGC GGGCACCGCC AGACCGGCCG GGGCCACCAT TGCCGCCGGG 480

CACCGAAACA ACAGCCCAAC GGTGCCGCCG GCCCCGCCGT TTGCCGCCAT CACCGGCCAT 540

TCACCGCCAG CACCGCCGTT AATGTTTATG AACCCGGTAC CGCCAGCGCG GCCCCTATTG 600

CCGGGCGCCG GAGNGCGTGC CCGCCGGCGC CGCCAACGCC CAAAAGCCCG GGGTTGCCAC 660

CGGCCCCGCC GGACCCACCG GTCCCGCCGA TCCCCCCGTT GCCGCCGGTG CCGCCGCCAT 720

TGGTGCTGCT GAAGCCGTTA GCGCCGGTTC CGCGGGTTCC GGCGGTGGCG CCNTGGCCGC 780

CGGCCCCGCC GTTGCCGTAT ACCCACCGCC CGTTGGCGTC GTTGCCGCCA TTGCCGCCAT 840

TGCCGCCGTT GCCGCCATTG CCGCCGTTCC CGTTGCCACC GCCGGNTTGG CCGCCGGCGC 900

CGCCGGCGGC CGC 913

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1872 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACATTGAAC GGCTTGATCC AGTTCGATGC CGCAATCCAG CCGGGTGATT CGGGCGGGCC 720

CGTCGTCAAC GGCCTAGGAC AGGTGGTCGG TATGAACACG GCCGCGTCCG ATAACTTCCA 780

GCTGTCCCAG GGTGGGCAGG GATTCGCCAT TCCGATCGGG CAGGCGATGG CGATCGCGGG 840

CCAAATCCGA TCGGGTGGGG GGTCACCCAC CGTTCATATC GGGCCTACCG CCTTCCTCGG 900

CTTGGGTGTT GTCGACAACA ACGGCAACGG CGCACGAGTC CAACGCGTGG TCGGAAGCCC 960

TCCGGCGGCA AGTCTCGGCA TCTCCACCGG CGACGTGATC ACCGCGGTCG ACGGCGCTCC 1020

GATCAACTCG GCCACCGCGA TGGCCGACGC GCTTAACGGG CATCATCCCG GTGACGTCAT 1080

CTCGGTGAAC TGGCAAACCA AGTCGGGCGG CACCGGTACA CGGAACGTGA CATTGGCCGA 1140

GGGACCGCCG GCCTGATTTG TGGCTGATAC CACCCTGCCG CCGGCCAACT GGATTGGCGC 1200

CACCGTGAT TGTCGCGTGA GCCCCCGAGT TCCGTCTCCC GTGCGCGTGG CATTGTGGAA 1260

GCAATGAACG AGGCAGAACA CAGCGTTGAG CACCCTCCGG TGCAGGGTAG TTACGTCGAA 1320

GGCGGTGTGG TCGAGCATCC GGATGCCAAG GACTTCGGCA GCGCCGCCGC GCTGCCGGCC 1380

GATCCGACCT GGTTTAAGCA CGCCGTCTTC TACGAGGTGC TGGTACTGGG CGTCTTCGAC 1440

GCCAGCGGCG ACGGTTCCGN CGATCTGCGT GGACTCATGG ATCGCCTCGA CTACCTGCAG 1500

TGGCTTGGCA TCGACTGCAT CTGTTGCCGC CGTTCCTACG ACTCACGCCT GCCCGACGGC 1560

CGTTACGACA TTCCCGACTT CTACAAGGTG CTGCCCGAAT TCGGCACCGT CGACGATTTC 1620

GTCGCCCTGG TCGACAAGGC TCACGAGCGA GGTATCAAGA TCATCACCGA CTTGGTGATG 1680

AATCACATGT CGGAGTCGCA CCCCTGGTTT CAGCAGTCAC GCTCGAGCAA AGACGACCG 124

TACGGTGACT ATTACGTCTG GAGTGACA AGGAGCTC AGAAGCCT AGATCGATC 18

TTCGTCAAG CTGAAGCT GCCGCCGT CGTCCTGCAG CCGCCGT 1

1

CTTCGCCGAA ACCTGATGCC GAGGAACAGG GTGTTCCCGT GAGCCCGACG GCGTCCGACC 60

CCGCGCTCCT CGCCGAGATC AGGCAGTCGC TTGATGCGAC AAAAGGGTTG ACCAGCGTGC 120

ACGTAGCGGT CCGAACAACC GGGAAAGTCG ACAGCTTGCT GGGTATTACC AGTGCCGATG 180

TCGACGTCCG GGCCAATCCC CTCCCGGCAA AGGGCGTATG CACCTACAAC GACGAGCAGG 240

GTGTCCCGTT TCGGGTACAA GCCGACAACA TCTCGGTGAA ACTGTTCGAC GACTGGAGCA 300

ATCTCGGGTC GATTTCTGAA CTGTCAACTT CACGCGTGCT CGATCCTGCC GCTGGCGTGA 360

TCGAGCTGCT GTCCCGTCTC ACGAACCTCC AAGCGCAAGG TACCGAAGTG ATACAGGGAA 420

TTTCGTCCAC CAAAATCACC GGGACGATTC CGGCAGTTC TGTCAAGATG CTTGATCCTG 480

GCGTCAAGAG TGCAAGGCCG GCGACCGTGT CGATTGCTCA GGACGGCTCG TACCAGCTCG 540

TCGGAACGAG CATCGACCTC GGATCCGGGT CGATTCAGCT CACGCAGTCG AAATGGAACG 600

AACCGATCAA CGTCGACTAG GCCGAAGTTG CGTCGACGCG TTTCTCGAAA TGCCCTTGTG 660

AACGGTGTCA ACGGCATCCG AAAACTGACC CCCTGACGGC ATCTGAAAAT TGACCCCCTA 720

GACCGGGCGG TTGGTGGTTA TTCTTCGGTG GTTCCGACTG GTGGGACGCG GCCGAGGTCG 780

CGGTCCTTGA GCCGGTAGCT GTCGCCTTTG AGGGCGACGA CTTCAGCATG GTGGACGAGG 840

CGGTCGATCA TGGCGGCAGC AACGACGTCG TCGCCGCCGA AAACCTCGCC CCACCGGCCC 900

AAGGGCTTAT TGGACGTGAC GATCAAGCTG GCCCGCTCAT ACGGGGAGGA CACCAGCTGG 960

AAGAAGAGGT TCGTGAGCTA CTCCTTAAAA CGAATGTAAA CGACTTCGTC AACGACCAGG 1020

ACGTGATAGC CGCCAAAACG CATGATTTCA TTAGATGC CGCCGCGTTC TGAGCGTCG 1080

GCAAATGCTT CTACTATTC GCTGAAACCA TTGACAAGA TTCAGAATT GGTTGACAC 1140

[illegible] 1200

[illegible] 1260

[illegible] 1320

[illegible] 1380

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATTCGGCA CGAGCCGGCG ATAGCTTCTG GGCCGCGGCC GACCAGATGG CTCGAGGGTT 60

CGTGCTCGGG GCCACCGCCG GGCGCACCAC CCTGACCGGT GAGGGCCTGC AACACGCCGA 120

CGGTCACTCG TTGCTCCTGG ACGCCACCAA CCCGGCGGTG GTTGCCTACG ACCCGGCCTT 180

CGCCTACGAA ATCGGTACA TCGNTGAAAG CGGACTGGCC AGGATGTGCG GGCAGAACCC 240

GGAGAACATC TTCTTCTACA TCACCGTCTA CAACGAGCCG TACGTGCAGC CGCCGGAGCC 300

GGAGAACTTC GATCCCGAGG GCGTGCTGGG GGGTATCTAC CGNTATCACG CGGCCACCGA 360

GCAACGCACC AACAAGGNGC AGATCCTGGC CTCCGGGGTA CCGATGCCCG CGGCGCTGCG 420

GGCAGCACAG ATGCTGGCCG CGGAGTGGGA TGTCGCCGCC GACGTGTGGT CGGTGACCAG 480

TTGGGGCGAG CTAAACCGCG ACGGCGTGGT CATCGAGACC GAGAAGCTCC GCCACCCCGA 540

TCGGCCGGCG GGCGTGCCCT ACGTGACGAG AGCGCTGGAG AATGCTCGGG GCCCGGTGAT 600

CGCGGTGTCG GACTGGATGC GCCCGGTCCC CGAGCAGATC CGACCGTGGG TGCCGGGCAC 660

ATACCTCACC TTGGGCACCG ACGGCTTCGG TTTTTCCGAC ACTCGGCCGC CGGTTGTCG 720

TTACTTCAAC ACCGACGCCG AACTCGAGTT TGTCGCCTT TTTCGGAGGG GTTTACCGGG 780

TCGACGGGTG AATACGGACC CATTCGGTGT CGTTGCTGG CCGCCGCCT ACTTACCGG 840

ATTTACAAA TTTCCCTTT TTTTTTTCC TAAGTT 876

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1471 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

CAGATTCATA ACGAATTCAC AGCGGCACAA CAATATGTCG CGATCGCGGT TTATTTCGAC 120

AGCGAAGACC TGCCGCAGTT GGCGAAGCAT TTTTACAGCC AAGCGGTCGA GGAACGAAAC 180

CATGCAATGA TGCTCGTGCA ACACCTGCTC GACCGCGACC TTCGTGTCGA AATTCCCGGC 240

GTAGACACGG TGCGAAACCA GTTCGACAGA CCCCGCGAGG CACTGGCGCT GGCGCTCGAT 300

CAGGAACGCA CAGTCACCGA CCAGGTCGGT CGGCTGACAG CGGTGGCCCG CGACGAGGGC 360

GATTTCGTCG GCGAGCAGTT CATGCAGTGG TTCTTGCACG AACAGATCGA AGAGCTGGCC 420

TTGATGGCAA CCCTGGTGCG GGTTGCCGAT CGGGCCGGGG CCAACCTGTT CGAGCTAGAG 480

AACTTCGTCG CACGTGAAGT GGATGTGCCG CCGGCCGCAT CAGGCGCCCC GCACGCTGCC 540

GGGGGCGCAC TCTAGATCCC TGGGCGGGAT CAGCGAGTGC TCCCGTTCGC CGGCCGGTCT 600

TCCAGCCAGG CCTTGGTGCC GCCGGGGTGG TGAGTACCAA TCCAGGCCAC CCCGACCTCC 660

CGGNAAAAGT CGATGTCCTC GTACTCATCG ACGTTCCAGG AGTACACCGC CCGGCCCTGA 720

GCTGCCGAGC GGTCAACGAG TTGCGGATAT TCCTTTAACG CAGGCAGTGA GGGTCCCACG 780

GCGGTTGGCC CGACCGCCGT GGCCGCACTG CTGGTCAGGT ATCGGGGGGT CTTGCCGAGC 840

AACAACGTCG GCAGGAGGGG TGGAGCCCGC CGGATCCGCA GACCGGGGGG GCGAAAACGA 900

CATCAACACC GCACGGGATC GATCTGCGGA GGGGGGTGCG GGAATACCGA ACCGGTGTAG 960

GAGCGCCAGC AGTTGTTTTC CCACCAGCGA AGGGTTTTCG GTTCATCGGN GGCCNTTAAG 1020

T 1021

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

[illegible]

TCCGCCCCTG ANGTCCTGAC CGCCGCCCAG TGGACCAGNC TGTTCAACAG NCTCGNCGAT 240

CCCAACGTGT CGTTTGNGAA CAAGGGNAGT CTGGTCGAGG GNGGNATCGG NGGNANCGAG 300

GGNGNGNATC GNCGANCACA A 321

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 373 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TCTTATCGGT TCCGGTTGGC GACGGGTTTT GGGNGCGGGT GGTTAACCCG CTCGGCCAGC 60

CGATCGACGG GCGCGGAGAC GTCGACTCCG ATACTCGGCG CGCGCTGGAG CTCCAGGCGC 120

CCTCGGTGGT GNACCGGCAA GCCGTGAAGG AGCCGTTGNA GACCGGGATC AAGGCGATTG 180

ACGCGATGAC CGGGATCGGC CGCGGGCAGG GCCAGCTGAT CATCGGGGAC CGCAAGACCG 240

GCAAAAACCG CGGTCTGTGT CGGACACCAT CCTCAAACCA GCGGGAAGAA CTGGGAGTCC 300

GGTTGATCCT AAGAAGCAAG TGCGCTTGTC TATACGTTGG TCATCGGGCA AGAAGGGGAA 360

CTTACCAGCG CGG 373

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

[illegible]

GCGTGGAGGT TTTCGTCACC GTCAGCCGTG GNAAGTGGGA CACGCTGCGC GCCATNGNGT 300

TTGACGACGA NCCATATCGG NGATTCCCNC ACATNCGAAG TTCCGANGGA GA 352

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAAATCGCGC TTCATTCCGT TCGACCAGCG GCTGGCGATA ATCGACGAAG TGATCAAGCC 60

GCGGTTCGCG GCGCTCATGG GTCACAGCGA GTAATCAGCA AGTTCTCTGG TATATCGCAC 120

CTAGCGTTCA GTTGCTTGCC AGATCGCTTT CGTACCGTCA TCGCATGTAC CGGTTCGCGT 180

GCGGCAGCCT CATGGTGGCT GCGTCGATCT TGCCACGGG TGTGGCGGGT CTCGGGGTCG 240

GCGCGTAGTC CGCAGCCCAA ACTCGCGCCG TGCCCGACTA CTACTGGTGC CCGGGGCAGC 300

CTTTCGACCC CGCATGGGGG CCCAACTGGG ATCCCTACAC CTGCCATGAC GACTTCCACC 360

CGGACAGCGA CGGCTCCGAC GACAACGGCT ACTACTCCGG ACCCATCCTC GAAGGTCCCG 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible]

[illegible]

[illegible]

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CGCGACGACG ACGAACGTCG GGCCCACCAC CGCCTATGCG TTGATGCAGG CGACCGGGAT 60

GGTCGCCGAC CATATCCAAG CATGCTGGGT GCCCACTGAG CGACCTTTTG ACCAGCCGGG 120

CTGCCCGATG GCGGCCCGGT GAAGTCATTG CGCCGGGGCT TGTGCACCTG ATGAACCCGA 180

ATAGGGAACA ATAGGGGTGT GATTTGGCAG TTCAATGTCG GGTATGGCTG GAAATCCAAT 240

GGCGGGGCAT GCTCGGCGCC GACCAGGCTC GCGCAGGCGG GCGAGCCCGA ATCTGGAGGG 300

AGCACTCAAT GGCGGCGATG AAGCCCCGGA CCGGCGACGG TCCTTTGGAA GCAACTAAGG 360

AGGGCGCGGG CATTGTGATG CGAGTACCAC TTGAGGGTGG CGGTCGCCTG GTCGTCGAGC 420

TGACACCCGA CGAAGCCGCC GCACTGGGTG ACGAACTCAA AGGCGTTACT AGCTAAGACC 480

AGCCCAACGG CGAATGGTCG GCGTTACGCG CACACCTTCC GGTAGATGTC CAGTGTCTGC 540

TCGGCGATGT ATGCCCAGGA GAACTCTTGG ATACAGCGCT 580

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH. 160 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

[illegible]

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible]
(C) [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GACACCGATA CGATGGTGAT GTACGCCAAC GTTGTCGACA CGCTCGAGGC GTTCACGATC 60

CAGCGCACAC CCGACGGCGT GACCATCGGC GATGCGGCCC CGTTCGCGGA GGCGGCTGCC 120

AAGGCGATGG GAATCGACAA GCTGCGGGTA ATTCATACCG GAATGGACCC CGTCGTCGCT 180

GAACGCGAAC AGTGGGACGA CGGCAACAAC ACGTTGGCGT TGGCGCCCGG TGTCGTTGTC 240

GCCTACGAGC GCAACGTACA GACCAACGCC CG 272

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCAGCCGGTG GTTCTCGGAC TATCTGCGCA CGGTGACGCA GCGCGACGTG CGCGAGCTGA 60

AGCGGATCGA GCAGACGGAT CGCCTGCCGC GGTTCATGCG CTACCTGGCC GCTATCACCG 120

CGTAGGAGCT GAACGTGCCC GAAGCGGCGC GGTTCATCGC GTTCGACGTG GAGACGATCC 180

GTTCGGATCT CGCGTTGGTC GAGATGGTCT ATCTGCTACA TCGACTGCTC GACTCGTCGC 240

GGAATCTGAC CGCGAAGATC AAGAAGCGGT CAAAGATCCA TTTCCTCGAC AGTCGCTTCG 300

CGCCCTGGTT CATCCGG 317

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

GCAGCGCCGG ACCACCTCGC CGGTGGTCAG CATGGTGATG ACCACGTCGG CGTCGGCGAC 120

CGCTTCGGGC GCGCTACGAA ACACCGCGAC ACCGTGCGCG GCGGCGCCGG ACGCCGCCGT 180

GG 182

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GATCGCGAAG TTTGGTGAGC AGGTGGTCGA CGCGAAAGTC TGGGCGCCTG CGAAGCGGGT 60

CGGCGTTCAC GAGGCGAAGA CACGCCTGTC CGAGCTGCTG CGGCTCGTCT ACGGCGGCCA 120

GAGGTTGAGA TTGCCCTGCG CGGCGAGCCG GTAGCAAAGC TTGTGCCGCT GCATCCTCAT 180

GAGACTCGGC GGTTAGGCAT TGACCATGGC GTGTACCCCG TGCCCGACGA TTTGGACGCT 240

CCGTTGTCAG ACCACGTGCT CGAACGCTTT CACCCGTGAA GCGCTACCTC ATCGACACCC 300

ACGTTTGG 308

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 267 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1539 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CTCGTGCCGA AAGAATGTGA GGGGACACGA TGAGCAATCA CACCTACCCA GTGATCGAGA 60

TCCTCGGGAC CTCGCCCCAC GGCGTCGACG CGGCAATCCA GGCCGCTCTG GCTCGAGCTG 120

CCTAGACTAT GCGCGCGCTG GACTGTTTCG AAGTACAGTC AATTCGAGGC CACCTGGTCG 180

ACGGAGCGGT CGCGCACTTC CAGGTGACTA TGAAAGTCGG CTTCCGCTGC AGCATTCCTG 240

AACCTTCAAG CGCGGCCGAT AACTGAGGTG CATCATTAAG CGACTTTTCC AGAACATCCT 300

GACGCGCTCG AAACGCCGTT CAGCCGACGG TGGCTCCGCC GAGGCGCTGC CTCCAAAATC 360

CCTGCGACAA TTCGTCGGCG GCGCCTACAA CGAAGTCGGT GCTGAATTCG TCGGGTATCT 420

GGTCGACCTG TGTGGGCTGC AGCCCGACGA AGCGGTGCTT GACGTCGGCT GCGGCTCGGG 480

GCCGATGGCG TTGCCGCTCA CCGGCTATCT GAACAGCGAG CGACGCTACG CCGGCTTCGA 540

TATCTCGCAC AAAGCCATCG CGTGGTCGCA GAACTACATC ACTTCGCGCG ATCGCGAACTT 600

CTCGTTCGAG GTCTTCGACA TTACAACTT CTTTACAAA CCGAAACCCA AATACTAGTC 660

ATTAGACTTT CGCTTTCCAT ACCTGGATGC CTCGCTCGAT GTGGTGTTTC TTACCTCGGT 720

TTTTATTAAA ATCTTTTTGC TTAATTTGAA TATTAATTTT AAAAAATTT TTTTTTGTAT 780

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 840

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 900

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 960

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1020

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1080

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1140

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1200

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1260

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1320

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1380

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1440

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1500

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAA 1539

AGCCTGCGTG ATCGGTCATC ACCAACGGTG ACAGCAGCCG GTTGTGCACC AGCGCGAACG 1320

CCACCCCGGT CTCCGGGTCT GTCCAGCCGA TCGAGCCGCC CAAGCCCACA TGACCAAACC 1380

CCGGCATCAC GTTGCCGATC GGCATACCGT GATAGCCAAG ATGAAAATTT AAGGGCACCA 1440

ATAGATTTCG ATCCGGCAGA ACTTGCCGTC GGTTGCGGGT CAGGCCCGTG ACCAGCTCCC 1500

GCGACAAGAA CCGTATGCCG TCGATCTCGC CTCGTGCCG 1539

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 851 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTGCAGGGTG GCGTGGATGA GCGTCACCGC GGGGCAGCCC GAGCTGACCG CCGCCCAGGT 60

CCGGGTTGCT GCGCCGGCCT ACGAGACGGC GTATCGCCTG ACGGTGCCCC CGCCGGTGAT 120

CGCCGAGAAC CGTCCTGAAC TGATGATTCT GATAGCGACC AACCTTTGCG GGCAAAACAC 180

CCCGGCGATC GCGCTGAACG AGGCCGAATA CGGCCAGATG TGGCCCCAAG ACGCCGCCGC 240

GATGTTTTGC TACCCGGTCG CGACCTCGAC GCTGAAGGCG ACGTTCCTGC CGTTCGAGGA 300

GGCGCTGGAG ATGACCAGCG CCTTTGGCCT GTTCGAGTAC GCGGTCGCGG TCGAGGAGGC 360

CTCGGACACC GTGCTGTTCA ACGACTTCAT GAAGAATTTC CGCTACGTCG TGAAAACGTT 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 660

[illegible] 720

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GATCGATCGG GCGGAAATTT GGACCAGATT CGCCTCCGGC GATAACCCAA TCAATCGAAC 60

CTAGATTTAT TCCGTCCAGG GGCCCGAGTA ATGGCTCGCA CGACAGGAAC CTTACTGCTG 120

CGGGCACCTG TCGTAGGTCC TCGATACCGC GGAAGCGCTC GACATTTTCC ACCGACACCC 180

CCATCCAAAC GTTCGAGGGC CACTCCAGCT TGTGACCGAG GCGACGCAGT CGCAGGCTGC 240

GCTTGGTCAA GATC 254

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1227 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CGCTGCTCAG CTTGGCCAAG GCCTGATCGG AGCGCTTGTC GCCGACGCCG TCGTCGATAC 600

CGCACAGCGC ATTGCGAACG ATGGTGTCCA CATCGCGGTT CTCCAGCGCG TTGAGGTATC 660

CCTGAATCGC GGTTTTGGCC GGTCCCTCCG AGAATGTGCC TGCCGTGTTG GCTCCGTTGG 720

TGCCGACCCC GTATATGATC GCCGCCGTCA TAGCCGACAC CAGCGCGAGG GCTACCACAA 780

TGCCGATCAG CAGCCGCTTG TGCCGTCGCT TCGGGTAGGA CAGCTGCCGC GGCACGCCGG 840

GATATGCGGC GGGCGCAGCC GCCGCGTCGT CTGCTCGTCC CGGGGCGAAG GCCGGTTCGG 900

CGGCGCCGAG GTCGTCGGGG TAGTCCAGGG CTTGGGGTTC GTGGGATGAG GGCTCGGCGT 960

ACGGCGCCGG TCGCTTCGTC CCACAGCTCG GTTCGGCCGA GTCGCGACCG GGCATTGTGG 1020

TTGTCCTAGG GTGGTGCACG GGAGCAGCTG CTAGGGCGAC AACCGCCCGT CGCGTCAGCC 1080

GGCAGCATCG GCAATCAGGT GAGCTCCCTA GGCAGGCTAG CGCAACAGCT GCCGTCAGCT 1140

CTCAACGCGA CGGGGCCGGC CGCGGCGCCG ATAATGTTGA AAGACTAGGC AACCTTAGGA 1200

ACGAAGGACG GAGATTTTGT GACGATC 1227

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 181 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GCGGTGTCGG CGGATCCGGC GGGTGGTTGA ACGGCAACGG CGGTGTCGGC GGCCGGGGCG 60

GCGACGGCGT CTTTGCCGGT GCCGGCGGCC AGGGCGGCCT CGGTGGGCAG GGCGGCAATG 120

GCGGCGGCTC CACCGGCGGC AACGGCGGTC TTGGCGGCGC GGGCGGTGGC GGAGGCAACG 180

GCCCCGGACGG CGGCTTCGGT GGCAACGGCG GTAAGGGTGG CCAGGGCGGN ATTGGCGGCG 240

GCACTCAGAG CGCGACCGGC GTCGGNGGTG ACGGCGGTGA CGGCGGTGAC 290

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GATCCAGTGG CATGGNGGGT GTCAGTGGAA GCAT 34

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible] acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

[illegible]

(A) LENGTH: 53 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATGGCGTTCA CGGGGCGGCG GGGACCGGGC AGCCCGGNGG GGCCGGGGGG TGG 53

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 132 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GATCCACCGC GGGTGCAGAC GGTGCCCGTG GTGCCACCTC GACCAGGCGC GGCAACGGCG 60

GCACCGGCGG CAACGGCGCG AACGCCACCG TCGTCGGNGG GGCCGGCGGG GCCGGCGGCA 120

AGGGTGGCAA CG 132

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 121 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCCGCCGCCG CCGCCGCCGC CGCCGCCGCC GCCGCCGCCG CCGCCGCCGC CGCCGCCGCC 60

GCCGCCGCCG CCGCCGCCGC CGCCGCCGCC GCCGCCGCCG CCGCCGCCGC CGCCGCCGCC 120

G 121

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 702 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTCGCCGGG TTTCGCCACC 60

CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGC GATCCGTTCG CGATGCCGGC 120

ATGAACGGGC GGCATCAAAT TAGTGCAGGA ACCTTTCAGT TTAGCGACGA TAATGGCTAT 180

AGCACTAAAG ACGATGATCC GATATGACAT ATCGCAGAC CGTGACGGTG GATCAGCAAG 240

AGATTTTGAA CAGGGCCAAC GAGGTGGAGC CCCCGATGGC GGACCCACCG ACTGATGTCC 300

CCATCACACC GTGCGAACTC ACGGNGGNTA AAAACGCCGC CGAACAGNTG GTNTTGTCCG 360

CCGACAACAT GCGCGAATAC CTGGCGGCCG GTGCCAAAGA GCGGCAGCGT CTGGCGACCT 420

CGCTGCGCAA CGCGGCCAAG GNGTATGCNG AGTTTGATGA CGACGCTGCG ACCGCCCTGG 480

ACAACGACGG CGAAGGAACT GTGCAGGCAG AAATCGCCGG GGCGTTCGGA GGGGACAGTT 540

CGGCCGAACT AACCGATACG CCGAGGGTCG CCATCGCCGG TGAACGCAAC TTCATGGATC 600

TCAAAGAAGC CGCAAGGAAG GTCGAAACCG CGGATAAGG CGCATCGCTC CGCTACTTGG 660

CGGATGTCTT GAACACTTNC ACCTGAACG CGAAGCGCAC GC 702

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

CTGGGCGGCG GTGGCATGAG AATGCCGATG GGTGCTCGCG ATCAGGGACA AGGGGGCGCC 240

AAGTCCAAGG GTTCTCAGCA GGAAGACGAG GCGCTCTACA CCGAGGATCC TCGTGCCG 298

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1058 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CGGCACGAGG ATCGAATCGC GTCGCGCGGA GCACAGCGTC GCACTGCACC AGTGGAGGAG 60

CCATGACCTA CTCGCCGGGT AACCCCCGAT ACCCGCAAGC GCAGCCCGCA GGCTCCTACG 120

GAGGCGTCAC ACCCTCGTTC GCCCACGCCG ATGAGGGTGC GAGCAAGCTA CCGATGTACC 180

TGAACATCGC GGTGGCAGTG CTCGGTCTGG CTGCGTACTT CGCCAGCTTC GGCCCAATGT 240

TCACCCTCAG TACCGAACTC GGGGGGGGTG ATGGCGCAGT GTCCGGTGAC ACTGCCCTGC 300

CGGTCGGCGT GGCTCTGCTG GTTGCGCTGC TTGCCGGGGT GGTTCTGGTG CCTAAGGCCA 360

AGAGCATGT GACGGTAGTT [illegible] 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 660

[illegible] 720

[illegible] 780

[illegible] 840

[illegible] 900

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TGGCACGAGA GACCGATGCC GCTACCCTCG CGCAGGAGCC AGGTAATTTC GAGCGGATCT 60

CCGGCGACCT GAAAACCCAG ATCGACCAGG TGGAGTCGAC GGCAGGTTCG TTGCAGGGCC 120

AGTGGCCGCG CGCGGCGGGG ACGGCGGCCC AGGTCGTGCT GATGGCGTTC CAAGAAGCAG 180

CCAATAAGCA GAAGCAGGAA CTCGACGAGA TCTTGACGAA TATTCGTCAG GCCGGCGTCC 240

AATACTCGAG GGCCGACGAG GAGCAGCAGC AGGCGCTATC CTCGCAAATG GGCTTCTGAC 300

CCGCTAATAC GAAAAGAAAC GGAGCAA 327

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 170 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATCCGGCGG CACGGGGGGT GCCGGCGGCA GCACCGCTGG CGCTGGCGGC AACGGCGGGG 60

CCGGGGGTGG CGGCGGAACC GGTGGGTTGC TCTTCGGCAA CGGCGGTGCC GGCGGGCACG 120

GGGCCGT 127

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CGGCGGCAAC GGGGGCACCG CCGGCAACGG GAGCGGCGCC GCCGGCGGCA ACGGCGGCAA 60

CGGCGGCTCC GGCCTCAACG G 81

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGGCACGAGA TCACACCTAC CGAGTGATCG AGATCGTCGG GACCTCGCCC GACGGTGTCG 60

ACCCCGNAAT CCAGGGCGGT CTGGCCCGAG CTGCTCAGAC CATGCCGCCC CTGGACTGGT 120

TCGAAGTACA GTCAATTCGA AGCTACTTGC TCGACGGAGC GGTCGCCGCA TTCCAGGTGA 180

CTATGAAAGT CGGTTTCCGC CTGGAGGATT CCTGAACGTT CAAGCGCGGC CGATAACTGA 240

GGTGCATCAT TAAGCGACTT TTCTAGAACA TCCTGATGCA CTCGAAACGC GGTTCAGCCG 300

ACGGTGGCTC CGCCGAGGCC CTGCCTCCAA AATCCTGCCA ATAATTCGTC GCCGG 355

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GCCAACGGGG TGTCTGGAAG CGCGTCGTAT TACGAAGTCA AGTTCAGCGA TCCGAGTAAG 720

CCGAACGGCC AGATCTGGAC GGGCGTAATC GGCTCGCCCG CGGCGAACGC ACCGGACGCC 780

GGGCCCCCTC AGCGCTGGTT TGTGGTATGG CTCGGGACCG CCAACAACCC GGTGGACAAG 840

GGCGCGGCCA AGGCCCTGGC CGAATCGATC CGGCCTTTGG TCGCCCCGCC GCCGGCGCCG 900

GCACCGGCTC CTGCAGAGCC CGCTCCGGCG CCGGCGCCGG CCGGGGAAGT CGCTCCTACC 960

CCGACGACAC CGACACCGCA GCGGACCTTA CCGGCCTGA 999

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 332 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Met His His His His His His Met His Gln Val Asp Pro Asn Leu Thr
1 5 10 15

Arg Arg Lys Gly Arg Leu Ala Ala Leu Ala Ile Ala Ala Met Ala Ser
20 25 30

Ala Ser Leu Val Thr Val Ala Val Ile Ala Thr Val Asn Asp Arg
35 40 45

Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Ser Pro Ser Thr
50 55 60

145 150 155 160

Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg
165 170 175

Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala
180 185 190

Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro
195 200 205

Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val
210 215 220

Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys
225 230 235 240

Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
245 250 255

Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
260 265 270

Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
275 280 285

Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro
290 295 300

Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Gly Gln Val Ala Pro Thr
305 310 315 320

Pro Thr Thr Pro Thr Pro Gln Asn Thr Leu Pro Ala
325 330

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys
1 5 10 15

Gln Gly Arg

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Glu Glu Ser Ile Ser Ile Xaa Xaa Xaa Ile Val Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Pro Lys Thr Tyr Xaa Xaa Xaa Xaa Leu Lys Gly Thr Arg Thr Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Asp Leu Ala Ser Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Ser
1 5 10 15

Leu Gln Ser Asn Asp Ser Asn Asp Gly Ile Asp Ile Glu Asp

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 187 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Thr Gly Ser Leu Asn His Thr His Asn Arg Arg Ala Asn Gln Arg Lys
1 5 10 15

Asn Thr Thr Met Lys Met Val Tyr Ser Ile Ala Asn Gly Leu Thr Ala
20 25 30

Asn Ala Ala Ala Leu Gly Ala Ala Asn Ala Gly Val Thr Ser Ile Met Ala
35 40 45

Gly Gly Pro Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro
50 55 60

Leu Pro Leu Asp Ile Ala Val Asn Pro Asp Val Pro Thr Ala Ala Gln
65 70 75 80

Leu Thr Ile Leu Leu Asn Val Leu Ala Asp Pro Asn Val Ser Phe Ala
85 90 95

Asn Lys Gly Ser Leu Val Ile Tyr Thr Leu Leu Gly Thr Ile Ala Arg
100 105 110

Thr Ala Arg Ser Gly Leu Lys Ala Ala Ser Thr Gly Asp Leu Pro
115 120 125

Ser Ile Ile Ser Thr Thr Asn Ala Ala Ser Thr Thr Asn
130 135 140

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Asp Glu Val Thr Val Glu Thr Thr Ser Val Phe Arg Ala Asp Phe Leu
1 5 10 15

Ser Ala Leu Asp Ala Ile Ala Thr Ala Gly Thr Ile Ser Ala Val Ser
20 25 30

Gly Val Glu Gly Leu Pro Phe Gly Ser Ala Leu Leu Val Val Lys Arg
35 40 45

Gly Ile Asn Ala Gly Ser Arg Phe Leu Leu Asp Ile Ala Thr Ser
50 55 60

Ala Gly Asn His Pro Asp Ser Asp Ile Phe Leu Asp Asp Val Thr Val
65 70 75 80

Ser Arg Arg His Ala Glu Phe Arg Leu Glu Asn Asn Glu Phe Asn Val
85 90 95

Val Asp Val Gly Ser Leu Asn Gly Thr Tyr Val Asn Arg Glu Pro Val
100 105 110

Asp Ser Ala Val Leu Ala Asn Gly Asp Glu Ile Gln Ile Gly Lys Leu
115 120 125

Arg Leu Ile Phe Leu Ser Glu Ala Pro Gln Thr Thr Gly Ile Ser
130 135 140

Ala Val Leu Val
145

(2) INFORMATION FOR SEQ ID NO:66:

Gly Pro Asp Arg Ser Ala Ser Leu Ser Leu Val Arg His Arg Arg Gln
20 25 30

Gln Arg Asp Ala Leu Cys Leu Ser Ser Thr Gln Ile Ser Arg Gln Ser
35 40 45

Asn Leu Pro Pro Ala Ala Gly Gly Ala Ala Asn Tyr Ser Arg Arg Asn
50 55 60

Ile Asp Val Arg Ile Lys Ile Leu Met Leu Val Thr Ala Val Val Leu
65 70 75 80

Leu Tyr Cys Leu Gly Val Ala Thr Ala Ala Phe Ile Lys Ile Tyr Cys Glu
85 90 95

Glu Leu Lys Gly Thr Asp Thr Gly His Ala Cys Glu Ile Gln Met Ser
100 105 110

Asp Pro Ala Ala Asn Ile Asn Ile Ala Leu Ser Leu Ser Tyr Tyr Asp
115 120 125

Gln Lys Ser Leu Asn Asn Tyr Leu Ala Gln Thr Arg Asp Lys Phe Leu
130 135 140

Leu Ala Asn Thr Ser Glu Pro Arg Glu Ala Ile Glu Arg Glu Leu Asn
145 150 155 160

Ile Thr Leu Arg Arg Thr Gln
165

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Thr Ala Ala Ser Asp Asn Phe Gln Leu Val Gln Gly Gly Gln Gly Phe
1 5 10 15

Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
20 25 30

Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
35 40 45

Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
50 55 60

Val Gly Ser Ala Pro Ala Asn Ser Ile Gly Ile Ser Thr Gly Asp Val
65 70 75 80

Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
85 90 95

Asp Ala Leu Asn Gly Arg Ser Gly Ser Val Val Ser Val Asn Trp
100 105 110

Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Gln
115 120 125

Gly Pro Pro Ala
130

INFORMATION FOR SEQ ID NO:7:

Ser Ala Gly Met Ala Arg Val Arg Arg Trp Xaa Val Met Pro Xaa Val
65 70 75 80

Ile Gln Ser Thr Xaa Ile Arg Xaa Xaa Gly Pro Phe Asp Asn Arg Gly
85 90 95

Ser Glu Arg Lys
100

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 168 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Met Thr Asp Asp Ile Leu Leu Ile Asp Thr Asp Gln Arg Val Arg Thr
1 5 10 15

Leu Thr Leu Asn Arg Pro Xaa Ser Arg Asn Ala Leu Ser Ala Ala Leu
20 25 30

Arg Asp Arg Leu Leu Ala Xaa Leu Xaa Asp Ala Glu Xaa Asp Asp Asp
35 40 45

Ile Asp Val Val Ile Leu Thr Gly Ala Gly Arg Ala Phe Val Ser Gly Ala Asp
50 55 60

Leu Asp Leu Leu Val Asn Leu Gln Thr Asp Arg Ala Ala Ala Gly Arg Phe
65 70 75 80

Ala Gln Leu Val Arg Ala Leu Leu Leu Leu Glu Val Pro Val Ile Ala Ala
85 90 95

Val Asn Gly Ala Ala Leu Gly Gly Gly Leu Glu Leu Ala Leu Ala Cys Asp
100 105 110

Leu Arg Val Ala Ala Ala Asp Ala Val Phe Ala Leu Pro Glu Val Arg Leu
115 120 125

Gly Leu Ile Pro Gly Ala Gly Gly Thr Gln Arg Leu Pro Arg Leu Val Gly
130 135 140

Val Ser Arg Ala Leu Glu Leu Ile Leu Thr Gly Arg Arg Leu Ser Ala Glu
145 150 155 160

Glu Ala Leu Arg Leu Gly Leu Val
165

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala Ala Leu Ser His His Leu Asp Thr Ala Pro His Leu Pro Pro Pro
225 230 235 240

Thr Arg Gln Val Val Arg Arg Val Val Gly Ser Trp His Gly Glu Pro
245 250 255

Met Pro Met Ser Ser Arg Trp Thr Asn Glu His Thr Ala Glu Leu Pro
260 265 270

Ala Asp Leu His Ala [illegible] Thr Arg Leu Ala Leu Leu Thr Gly Leu Ala
275 280 285

Pro His Gln Val Thr Arg Arg Asp Val Ala Ala Ala Arg Ser Leu Leu
290 295 300

Asp Thr Asn Ala Ala Leu Val Gly Ala Leu Ala Trp Ala Ala Pro Thr
305 310 315 320

Ala Ala Arg Arg Ile Gly Thr Trp Leu Gly Ala Ala Ala Glu Gly Gln
325 330 335

Val Val Arg Gln Asn Pro Thr Gly
340

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 495 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

Arg Glu Arg Tyr Leu Leu His Asp Gln Gln Leu Arg Ile Ala Gln Ser
85 90 95

Thr Gly Glu Leu Met Asp Arg Ser Ala Arg Cys Val Ala Ala Ala Glu
100 105 110

Asp Gln Tyr Glu Pro Gly Ser Ser Arg Arg Trp Ala Glu Arg Phe Ala
115 120 125

Thr Leu Leu Arg Asn Leu Gln Phe Leu Pro Asn Ser Pro Thr Leu Met
130 135 140

Asn Ser Gly Thr Asp Leu Gly Leu Leu Ala Gly Cys Phe Val Leu Pro
145 150 155 160

Ile Gln Asp Ser Leu Gln Ser Ile Ile Ala Thr Leu Gly Gln Ala Ala
165 170 175

Leu Leu Leu Arg Asn Gly Tyr Val Thr Gly Tyr Asn Ile Ser His Leu
180 185 190

Asn Pro Ala Gly Asp Arg Val Ala Ser Thr Gly Gly Thr Ala Ser Gly
195 200 205

Ile Val Ser Ile Leu Arg Leu Tyr Arg Ser Ala Ala Gly Thr Val Ser
210 215 220

Met Gly Gly Arg Asn Ala Gln Ala Tyr Met Ala Val Leu Asp Val Ser
225 230 235 240

Ile Ser Asn Leu Tyr Asn Ile Thr Ala Asp Asn Leu Ser Ile Ser
245 250 255

Ile Ile Asn Ile Gln Ser Ala Thr Ile Ser Ala Ile Ser
260 265 270

Asn Ile Val Ile Asn Asn Thr Ile Ser Ala Ile Ser Ser
275 280 285

370 375 380

Asp Val Ile Asp Val Ser Arg Tyr Pro Phe Pro Glu Leu Gly Glu Ala
385 390 395 400

Ala Arg Ala Thr Arg Lys Ile Gly Leu Gly Val Met Gly Leu Ala Glu
405 410 415

Leu Leu Ala Ala Leu Gly Ile Pro Tyr Asp Ser Glu Glu Ala Val Arg
420 425 430

Leu Ala Thr Arg Leu Met Arg Arg Ile Gln Gln Ala Ala His Thr Ala
435 440 445

Ser Arg Arg Leu Ala Gln Ile Arg Gly Ala Ile Pro Ala Phe Thr Asp
450 455 460

Ser Arg Phe Ala Arg Ser Gly Pro Arg Arg Arg Ala Glu Val Thr Ser
465 470 475 480

Val Ala Pro Thr Gly
485

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

[illegible]

Thr Pro Thr Ala Ala Val Ser Pro Pro Val Leu Lys Glu Gly Asp
110

Asp Cys Pro Asp Ser Thr Leu Ala Val Lys Gly Leu Thr Asn Ala Pro
115 120 125

Gln Tyr Tyr Val Gly Asp Gln Pro Lys Phe Thr Met Val Val Thr Asn
130 135 140

Ile Gly Leu Val Ser Lys Lys Arg Asp Val Gly Ala Ala Val Leu Ala
145 150 155 160

Ala Tyr Val Tyr Ser Leu Asp Asn Lys Arg Leu Trp Ser Asn Leu Asp
165 170 175

Cys Ala Pro Ser Asn Cys Thr Leu Val Lys Thr Ser Ser Pro Gly Glu
180 185 190

Cys Val Thr Thr Ala Val Thr Trp Pro Gly Ser Gly Asn Ala Pro Arg
195 200 205

Ser Pro Leu Pro Arg Pro Ala Ile Gly Leu Gly Thr Tyr Asn Leu Val
210 215 220

Val Ser Leu Gly Asn Leu Arg Ser Leu Val Pro Phe Ile Leu Asn
225 230 235 240

Gln Pro Pro Pro Pro Pro Gly Leu Val Ala Pro Gly Leu Ala Gln
245 250 255

Ala Pro Pro Leu Gln Ala Gly Gly
260

Lys Val Asp Asp Arg Pr Ile Asn Ser Ala Asp Ala Leu Val Ala Ala
50 55 60

Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr Phe Gln Asp
65 70 75 80

Pro Ser Gly Gly Ser Arg Thr Val Gln Val Thr Leu Gly Lys Ala Glu
85 90 95

Gln

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Gly Ala Ala Val Ser Leu Leu Ala Ala Gly Thr Leu Val Leu Thr Ala
1 5 10 15

Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser
20 25 30

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
35 40 45

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
50 55 60

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
65 70 75 80

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
85 90 95

Ala Ala Ala Ala
100

145 150 155 160

Gln Ile Gln Ala Leu Asn Ser Gly Thr Asp Leu Pro Ile Thr Pro Ile
165 170 175

Ser Val Ile Phe Arg Ser Asp Lys Ser Gly Thr Ser Asp Asn Phe Gln
180 185 190

Lys Tyr Leu Asp Gly Val Ser Asn Gly Ala Trp Gly Lys Gly Ala Ser
195 200 205

Glu Thr Phe Ser Gly Gly Val Gly Val Gly Ala Pro Gly Asn Asn Gly
210 215 220

Thr Ser Ala Leu Leu Gln Thr Thr Asp Gly Ser Ile Ile Tyr Asn Gln
225 230 235 240

Trp Ser Phe Arg Val Gly Lys Gln Leu Asn Met Asn Gln Ile Ile Thr
245 250 255

Ser Ala Gly Pro Asp Pro Val Ala Ile Ile Thr Thr Glu Ser Val Gly Lys
260 265 270

Thr Ile Ala Gly Ala Lys Ile Met Gly Gln Gly Asn Asp Leu Val Leu
275 280 285

Asp Thr Ser Ser Phe Tyr Arg Pro Thr Ile Pro Gly Ser Tyr Pro Ile
290 295 300

Val Leu Ala Thr Tyr [illegible]
305 310 315 320

[illegible]
325 330 335

[illegible]
340 345 350

Gln Ala Ala Ala Gly Arg Ala Val Arg Arg Thr Gly His Ala Glu Asp
1 5 10 15

Gln Thr His Gln Asp Arg Leu His His Gly Cys Arg Arg Ala Ala Val
20 25 30

Val Val Arg Gln Asp Arg Ala Ser Val Ser Ala Thr Ser Ala Arg Pro
35 40 45

Pro Arg Arg His Pro Ala Gln Gly His Arg Arg Arg Val Ala Pro Ser
50 55 60

Gly Gly Arg Arg Arg Pro His Pro His His Val Glu Pro Asp Asp Arg
65 70 75 80

Arg Asp Arg Pro Ala Leu Pro Asp Arg Thr Gln Pro Ala Glu His Pro
85 90 95

Asp Pro His Arg Arg Gly Pro Ala Asp Pro Gly Arg Val Arg Gly Arg
100 105 110

Gly Arg Leu Arg Arg Val Asp Arg Gly Arg Leu Gln Pro Asp Arg Asp
115 120 125

Ala Asp His Gly Ala Pro Val Arg Gly Arg Gly Pro His Arg Gly Val
130 135 140

Gln His Arg Gly Gly Pro Val Pro Val Arg Arg Val Pro Gly Val Arg
145 150 155 160

Cys Ala His Arg Asp Gly His Arg Arg Val Ala Ala Pro Gly Gln Gly
165 170 175

Arg Val Leu Arg Arg Asp Gly Pro Arg Leu Arg Pro Ala
180 185 190

Ala Val Glu Arg Leu His Arg Gly Val Ser Arg Ala Asp Gly Arg Val
195 200 205

290 295

Asn Arg Pro Arg Arg
305

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 580 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Met Ala Val Trp Cys Leu Asn Gly Leu Thr Gly Arg His Arg His Gly
1 5 10 15

Arg Cys Arg Val Arg Ala Met Gly Thr Arg Ser Ser Asn Arg Trp Cys
20 25 30

Ser Thr Thr Ala Asp Tyr Leu Ala Ser Leu Thr Pro Thr Glu Ala Ala
35 40 45

Ser Pro Leu Met Arg Arg Phe Thr Tyr Tyr Ser Ile Ala Val Gly Cys
50 55 60

Arg Ile Arg Leu Phe Ile Val Arg Arg Leu Ala Met Gly Arg Ile Ile
65 70 75 80

Cys Arg Thr Leu Gly Val Arg Arg Ile Leu Leu Arg Thr Leu Ala Ile
85 90 95

Ile Arg Ala Glu Ile Leu Leu Ala Thr Met Cys Leu Leu Ala Ile His
100 105 110

Gly Leu Ile Pro Gly Val Ile Pro Thr Met Thr Pro Pro Pro Gly Met
195 200 205

Val Arg Gln Arg Pro Arg Ala Gly Met Leu Ala Ile Gly Ala Val Thr
210 215 220

Ile Ala Val Val Ser Ala Gly Ile Gly Gly Ala Ala Ala Ser Leu Val
225 230 235 240

Gly Phe Asn Arg Ala Pro Ala Gly Pro Ser Gly Gly Pro Val Ala Ala
245 250 255

Ser Ala Ala Pro Ser Ile Pro Ala Ala Asn Met Pro Pro Gly Ser Val
260 265 270

Glu Glu Val Ala Ala Ser Val Val Pro Ser Val Val Met Leu Ser Thr
275 280 285

Asp Leu Gly Arg Ser Ser Glu Pro Gly Ser Gly Ile Ile Leu Ser Ala
290 295 300

Glu Gly Leu Ile Leu Thr Asn Asn His Val Ile Ala Ala Ala Ala Lys
305 310 315 320

Pro Pro Leu Gly Ser Pro Ile Ile Lys Thr Thr Val Thr Ile Ser Asp
325 330 335

Gly Arg Thr Ala Ile Phe Thr Val Val Gly Ala Asp Ile Thr Ser Asp
340 345 350

Val Ala Val Leu Arg Val Gln Gly Val Ser Gly Leu Thr Pro Ile Ser
355 360 365

Ile Gly Asp Ser Asp Ser Leu Arg Val Gly Asp Pro Val Ile Ala Ile
370 375 380

Gly Asn Pro Leu Gly Leu Gly Gly Thr Val Thr Thr Gly Ile Val Ser
385 390 395 400

485 490 495

Val Gln Val Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Glu
500 505 510

Val Val Ala Gly Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val
515 520 525

Val Val Thr Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu
530 535 540

Val Ala Ala Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr
545 550 555 560

Thr Gln Asp Ile Ser Gly Gly Ser Ala Thr Val Gln Val Thr Leu Gly
565 570 575

Pro Ala Glu Gln
580

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 235 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Val Gln Gly Asp Asn Ile Ser Val Lys Leu Ile Asp Asp Trp Ser Asn
115 120 125

Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala
130 135 140

Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn Leu Gln Ala Gln
145 150 155 160

Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys Ile Thr Gly Thr
165 170 175

Leu Pro Ala Ser Ser Val Lys Ala Leu Asp Pro Gly Ala Lys Ser Ala
180 185 190

Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser His His Leu Val
195 200 205

Arg Ala Ser Leu Arg Leu Lys Ser Gly Ile Gln Leu Thr Gln Ser
210 215 220

Lys Trp Asn Glu Pro Val Asn Val Asp
225 230

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

(ii) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Val Pro Pro Ala Pro Pr[illegible] Leu Pro Pro Leu Pro Pro Ser Pro Ile Ser
1 5 10 15

Tyr Ala Ser Pr[illegible] Pro Ser Pr[illegible] Pro Leu Pro Tyr Asp Pro Pro Val Ala
20 25 26

Pro Gly Pro Pr[illegible] Met Pr[illegible] Pr[illegible] Leu Asp Pro Trp Pro Pro Ala Pro Pro
35 40 45

Leu Pr[illegible] Tyr Ser Thr Pro Pro Gly Ala Pro Leu Pro Pro Ser Pro Pro
50 55 60

Ser Pro Pro Leu Pro
65

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

Gly Ile Val Ile Asp Ile Asn Gly Val Val Leu Thr Asn Asn His Val
85 90 95

Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln
100 105 110

Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala
115 120 125

Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly
130 135 140

Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly
145 150 155 160

Gly Gln Gly Gly Glu Pro Arg Ala Val Pro Tyr Arg Val Val Ala Leu
165 170 175

Gly Gln Thr Val Glu Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu Thr
180 185 190

Leu Asn Gly Leu Ile Gln Thr Asp Ala Ala Ile Glu Pro Gly Asp Ser
195 200 205

Gly Gly Pro Val Val Asn Gly Ser Gln Val Val Gly Met Asn Thr
210 215 220

Ala Ala Ser Asp Asn Phe Gln Leu Ser Glu Gly Tyr Gln Gly Phe Ala
225 230 235 240

Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Glu Gln Ile Arg Ser Gly
245 250 255

Gly Val Val Asp Asn Ala Ala Arg Ala Glu Ile Gly Val Ser Ile Gln
260 265 270

Asp Val Thr Pro Asp Leu Ala Glu Ser Leu Gly Leu Asp Arg Pro Arg
275 280 285

Gly Ala Leu Val Ser Gln Val Leu Pro Asn Ser Pro Ala Ala Lys Ala
290 295 300

Gly Leu Gln Ala Gly Asp Val Ile Leu Ser Val Asn Gly Gln Ser Val
305 310 315 320

Val Ser Ala Asn Asp Val Ser Asp Val Val Lys Arg Glu Ser Thr Leu
325 330 335

Asn Met Val Val Leu Arg Asp Gly Arg Arg Met Thr Val Asn Leu Gln
340 345 350

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 205 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Ser Pro Lys Pro Asp Ala Glu Gln Gln Glu Gly Val Ile Val Pro Thr
1 5 10 15

Ala Ser Asp Pro Ala Leu Leu Ala Glu Leu Arg Glu Glu Asp Ala
20 25 30

Thr Lys Gly Leu Thr Ser Val His Val Ala Val Ser Ile Leu Gly Lys
35 40 45

Val Asp Ser Leu Leu Gly Ile Thr Leu Arg Asp Val Asp Val Arg Ala
50 55 60

Asn Pro Leu Ala Ala Lys Gly Val Lys Ser Tyr Ser Asp Val His Gly
65 70 75 80

Val Pro Pro Arg Val Glu Gly Arg Asn Leu Ile Val Thr Ile Asp
85 90 95

Val Ser Ser Arg Leu Gly Ser Val Thr Leu Ser Thr Val Leu Ala Arg
100 105 110

Ala Ser Ala Ala Gly Val Ile Ala Val Thr Asp
115 120

[illegible]

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Asn Asn Gly Phe Val
1 5 10 15

Leu Gly Ala Thr Asn Gly Arg Thr Thr Leu Thr Gly Thr Gly Leu Gln
20 25 30

His Ala Asp Gly His Ser Leu Leu Leu Asp Asn Thr Asn Pro Asn Val
35 40 45

Val Ala Tyr Asp Pro Ala His Ala Tyr Gln Ile Gly Tyr Thr Xaa Glu
50 55 60

Ser Gly Leu Ala Asn Met Tyr Gly Thr Asn Pro Glu Asn Ile Phe Phe
65 70 75 80

Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Glu Pro Glu
85 90 95

Asn Ile Asp Pro Ser Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala
100 105 110

Ala Thr Pro Gln Asn Thr Asn Pro Xaa Xaa Ile Ile Ala Ser Val Val
115 120 125

Asn Met Leu Ala Xaa Xaa Asn Met Asn Asn Ala Ile Thr
130 135 140

225 230 235 240

Ala Gln Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg
245 250 255

Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln
260 265 270

Leu Pro Gly Phe Asp Glu Gly Gly Gly Leu Arg Pro Xaa Lys
275 280 285

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

Thr Lys Phe His Ala Leu Met [illegible] [illegible] Gln Ile His Asn Gln Phe Thr
1 5 10 15

Ala Ala [illegible] Asp

[illegible] Ala Ser Gly Asn Pro His [illegible] Ala Gly Gly Arg [illegible]
[illegible] [illegible]

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Asn Ala Asp [illegible] Lys Asn [illegible] Thr Met Tyr Met Val Tyr [illegible] Ile
1 [illegible] [illegible] [illegible]

Asn Ala Gly Leu Thr Ala Ala Ala Ala Ile Gly Ala Asn Ala Ala Gly
20 25 30

Val Thr Ser Ile Met Ala Gly Gly Pro Thr Val Tyr Gln Met Gln Pro
[illegible] [illegible] 45

Val Val Ile Gly Asn Pro [illegible] Leu Asp Pro Xaa Ser Ala Pro Xaa
[illegible] [illegible] 60

Val Pro Thr [illegible] Gln Trp [illegible] Xaa [illegible] Asn Xaa Leu Xaa Asp
[illegible] [illegible] [illegible] [illegible]

[illegible]

[illegible]

[illegible]

Arg Asn Ala Leu Glu Leu Gln Ala Pro Ser Val Val Xaa Asp Gln Gly
35 40 45

Val Lys Glu Pro Leu Xaa Thr Gly Ile Lys Ala Ile Asp Ala Met Thr
50 55 60

Pro Ile Gly Arg Gly Gln Arg Gln Leu Ile Ile Gly Asp Arg Lys Thr
65 70 75 80

Gly Lys Asn Ala Ile Ala Leu Asp Ala Thr Pro Ser Ser Asn Gln Arg Glu
85 90 95

Ala Leu Gly Val Arg Trp Ile Pro Asn Ser Arg Cys Ala Ala Cys Val Tyr
100 105 110

Val Gly Glu Arg Ala Ala Glu Val Gln Glu Phe Leu Arg Asp
115 120 125

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Arg Ser Ser Arg Gly
115

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 103 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Met Tyr Arg Ile Asn Tyr Arg Thr Leu Met Leu Asn Asn Tyr Ile Leu
1 5 10 15

Ala Thr Gly Tyr Ala Gly Leu Gly Val Gly Ala Thr Ser Asn Ala Ala
20 25 30

Thr Ala Pro Val Pro Asp Tyr Tyr Trp Tyr Pro Gly Asn Ile Ile Asp
35 40 45

Pro Ala Thr Gly Pro Asn Ile Asp Pro Tyr Thr Tyr His Asp Asp Phe
50 55 60

His Asn Asp Ser Ser Asp Arg Asp Arg Ile Ile Ala Asn Ser Gly Pro
65 70 75 80

Ile Ser Ser Thr Gly Ile Gly Asn Ala Ser Asn Ala Ala Ile Pro
85 90 95

Arg Ser Ser Arg Gly
100

Ala Asp Glu Ala Arg Asn Gly Gly Ile Ala Arg Ile Trp Arg Glu His
30

Ser Met Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala
35 40 45

Thr Lys Glu Gly Arg Gly Ile Val Met Arg Val Pro Leu Glu Gly Gly
50 55 60

Gly Arg Leu Val Val Gly Leu Thr Pro Asp Glu Ala Ala Ala Leu Gly
65 70 80

Arg Thr Leu Leu Lys Lys Val Thr Pro
85

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 95 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Thr Asp Ala Arg Thr Leu Arg Gln Thr Ala Trp Arg Ile Leu Arg Ile
15

Arg Asp Ile Ala Thr Ile Asp Arg Thr Ala Gly

Ser Val Thr Ala Ala Ser Gly Ala Arg Ala Asn Ala Ala Arg

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
1 5 10 15

Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Thr Asp Val
20 25 30

[illegible]
35 40 45

[illegible]
50 55 60

[illegible]
65 70 75 80

[illegible]
85 90 95

[illegible]
100 105 110

[illegible]
115 120 125

[illegible]
130 135 140

[illegible]
145 150 155 160

[illegible]
165 170 175

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 263 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala
225 230 235 240

Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly
245 250 255

Arg Arg Asn Gly Gly Pro Ala
260

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 403 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Met Thr Tyr Ser Xaa Xaa Asn Xaa Xaa Tyr Xaa Xaa Ala Xaa Xaa Ala
1 5 10 15

Gly Ser Tyr Arg Gly Val Xaa Xaa Xaa Ser Xaa Ala His Ala Asp Glu Gly
20 25 30

Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly
35 40 45

Leu Ala Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Met Xaa Xaa Xaa Xaa Ser Thr
50 55 60

Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

Ala Ile Ala Pro Arg Pro Lys Ile Asn Pro Tyr Gly Gln Tyr Gly Arg
165 170 175

Tyr Gly Gln Tyr Gly Gln Tyr Gly Val Gln Pro Gly Gly Tyr Tyr Gly
180 185 190

Gln Gln Gly Ala Gln Gln Ala Ala Gly Leu Gln Ser Pro Gly Pro Gln
195 200 205

Gln Ser Ile Gln Pro Pro Gly Tyr Gly Ser Gln Tyr Gly Gly Tyr Ser
210 215 220

Ser Ser Ile Ser Gln Ser Ser Gly Ser Gly Tyr Thr Ala Gln Pro Pro Ala
225 230 235 240

Gln Pro Pro Ala Gln Ser Gly Ser Gln Gln Ser His Gln Gly Pro Ser
245 250 255

Ala Ile Pro Thr Gly Phe Ser Ile Ser Pro Pro Pro Pro Val Ser
260 265 270

Ala Gly Thr Gly Ser Gln Ala Gly Ser Ala Ile Val Asn Tyr Ser Asn
275 280 285

Pro Ser Gly Gly Gln Gln Val Ser Ser Ile Gly Gly Ala Ile Val
290 295 300

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

[illegible] 480

[illegible] 507

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 168 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

[illegible]

(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

[illegible] 500

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible] acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

Glu Asn Ala Ala Met Ala Ala Gln Leu Lys Ala Val Val Gly Ala Ala
65 70 75 80

Gln Tyr Ile Gly Leu Val Glu Ser Val Ala Gly Ser Cys Asn Asn Tyr
85 90 95

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

ACTATAAAAT AGTATCTAAA TTTTGCTTTT ATTAATAATC TTCTAAGCTC AATCTTAGGGA 60

AATGTCACTT CTATTTATTC TTTCTTTTAT TATTTAAATC ATTCTCTGAT TAAGCTCGCA 120

TTTCTTTTTT CTCTAGTCTT TTCTGAATCT TATTTT 156

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

[illegible]

[illegible]

TGATCGCCGA GAACCGTGCT GAACTGATGA TTCTGATAGC GACCAACCTC TTGGGGCAAA 780

ACACCCCGGC GATCGCGGTC AACGAGGCCG AATACGGCGA GATGTGGGCC CAAGACGCCG 840

CGGCGATGTT TGGCTACGCC GCGGCGACGG CGACGGCGAC GGCGACGTTG GTAGTCGTTC 900

AGGAGGGCGC GGAGATGACC AGTCCGTTG GTTTTTTA GTAGCCGCAT TTGGTGAGG 960

A[illegible] 1020

[illegible] 1080

[illegible] 1140

[illegible] 1200

[illegible] 1260

[illegible] 1320

[illegible] 1380

[illegible] 1440

[illegible] 1500

[illegible] 1560

[illegible] 1620

[illegible] 1680

GCCGGCAACA ACATGGCGCA AATCGACAGC GTCGTTGGTT CGAGTTGGGA CTGACAGCA 2340

GCCAAGGCCA GGGACGTGGT TTACGAGTGA AGTTCGTCGC GTGATCGTTC GGGTGGCAGT 2400

CTAAGTGGTC AGTGCTGGGG TGTTGGTGGT TTGCTGCTTG GCGGGTTCTT CGGTGCTGGT 2460

CAGTGCTGCT CGGGCTCGGG TGAGGACGTC GAGGCCCAGG TAGCGCCGTC CTTCGATCCA 2520

TTCGTCGTGT TGTTCGGCGA AGACGGTTCC GACGAGGCGG ATGATCGAGG CGTGGTCGGG 2580

GAAGATGCCC ACGACGTTGG TTCGCGTCCT TACCTCTCGC TTGAGGCGTT CCTGGGGGTT 2640

CTCGACGTAG ATTTCGCCCT AGATCTCGTT GGCGAAGGCG GTGAACGCCA GCAGGTCGGT 2700

CGCGCGGTTG TCGACCTGTT CGTCGACCTG GGGATTTTG TCGGTTAGAG CGTCGAGTAC 2760

CTGATCATAT TCGCAACAA CTGATTCGGC GTCGCTTTCC TCGTAGATGC ACTTCGCGCG 2820

GTTGCGCACC CACGGCCAGG AGGGCTTCGG GGTGCTGCTC ATCAGATTGG TCTTTAGTG 2880

GTTTCTGCAG CGCTGCCAGG CGGCTGCCGG CACTTCGCCC CGGATTGCGG CTACGAGGCC 2940

CCGGTGGGCG TCGCTGCTGA CAACCGCAAC CTTGGACAGC CGGCGGGCGA CGACGTCGCC 3000

GAGGAAGCCC ACGCAGTCGA CGTCGTCGTA GTCGAGCTTG ACTTGGATGG CTACGATC 3058

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino
(B) TYPE: amino acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Asn Val Ala Ala Ala
85 90 95

Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
100 105 110

Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly
115 120 125

Gln Asn Thr Pro Ala Ile Ala Val Asn Gln Ala Glu Tyr Gly Glu Met
130 135 140

Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala
145 150 155 160

Thr Ala Thr Ala Thr Leu Leu Ile Phe Gln Glu Ala Ile Thr Met Thr
165 170 175

Ser Ala Gly Gly Ser Leu Gln Ile Ala Ala Ala Val Gln Gln Ala Ser
180 185 190

Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu
195 200 205

Gln Gln Leu Ala Gln Ile Thr Ile Gly Thr Leu Pro Ala Asp Tyr Leu
210 215 220

Gly Gly Leu Trp Tyr Ile Val Ile Ile Gly Ala Ser Ile Ile Asn Asn
225 230 235 240

Val Val Ser Met Ala Ala Asn Met Met Ser Met Thr Asn Asn Gly Val
245 250 255

Ala Met Thr Asn Thr Val Ile Ile Met Ile Gly Phe Ala Ile
260 265 270

Xaa Xaa Leu Ser Gly Val Leu Arg Val Ile Pro Arg Pro Tyr Val Met
370 375 380

Pro His Ser Pro Ala Ala Gly
385 390

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1725 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

GAGGTAATA CTTTCGTTT AGGTTTGGAA CTTGATTGAT TTTGATCTGT TTTCAAGTTG 60

ATGTCCTTGG CTGTTTGGT GGCGTGGATG TAGATTTAT TCTTTCTTT ATTGCAACTA 120

ATTTTTTTA ATTTTTGTT ATGTATATTA TTTATTATT GTTAATTTA GTTTCATGG 180

CTTTCGTTT TTATTTTTT TTTATTATT TTATTAAAT TTTTGTTGT TGTTAGGCG 240

TTTATTAGA TTTTTTATT ATTAATTAA TTGAATTTT TGTTGTTAT TTTGGTTAA 300

ATTATATTA TATTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT 360

TTAATTTTT TTTTTTTTT TATTTTTTT TTTTTTTTT TTTTTTTAT TTTTTTTAT 420

TTTTTTTTT TTTTTTTTT TTTTATTTT TTTTTTTTA TTAATTTTT 480

TATTTATTT TTTTTTTTA TTTTTTTTT TTTTTTTAT TTTTTTTTT 540

[illegible] 1140

GGCGTTGCTG CGGTTCGAGC ACGCCGCACT GATTACCAAC CCCGGCGGGC TCCTTGAGCA 1200

GGCCGTCGCG GTCGACGAGC CCATCGACAC CGCCCCGGCG AACCAGTTGA TGAACAATGT 1260

GCCCCAAGCG CTGCAACAGC TGGCCCAGCC AGCGCAGGGC GTCGTACCTT CTTCCAAGCT 1320

GGGCGGGCTG TGGACGTCGG TCTGGACGCA TCTGTCGCCG CTCAGCAACG TCAGTTCGAT 1380

ATCCAACTAC CACATCTTA TGATGGCA [illegible] ATGACGAACA GCTTGCACTC 1440

[illegible] 1500

[illegible] 1560

[illegible] 1620

[illegible] 1680

CAGCCTGACC AGCCGCTCGC AAACGCGCCT TGATACATC CTGCG 1725

(2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) STRANDEDNESS:
(D) TOPOLOGY: [illegible]

Ala Tyr Leu Thr Ala Tyr Asn Leu Ile Val Ile Pro Ile Val Ile Ala
110 115 120

Glu Asn Arg Thr Glu Leu Met Thr Leu Thr Ala Thr Asn Leu Leu Gly
115 120 125

Gln Asn Thr Pro Ala Ile Glu Ala Asn Gln Ala Ala Tyr Ser Gln Met
130 135 140

Trp Gly Gln Asp Ala Glu Ala Met Tyr Gly Tyr Ala Ala Thr Ala Ala
145 150 155 160

Thr Ala Thr Leu Ala Leu Leu Pro Phe Ile Asp Ala Pro Leu Ile Thr
165 170 175

Asn Ile Gly Tyr Leu Ile Ile Leu Val Ala Val Ala Val Gly Thr Ala Ile
180 185 190

Asp Thr Ala Ala Asn Asn Ile Leu Met Asn Asn Val Ile Val Asn Leu
195 200 205

Gln Gln Leu Ala Gln Pro Ala Gln Tyr Val Val Ile Pro Pro Tyr Leu
210 215 220

Lys Gly Leu Trp Thr Ala Val Leu Leu Met Leu Ser Ile Leu Ser Asn
225 230 235 240

Thr Ser Ser Ile Ala Asn Asn Ile Met Val Met Gly Thr Gly Val
245 250 255

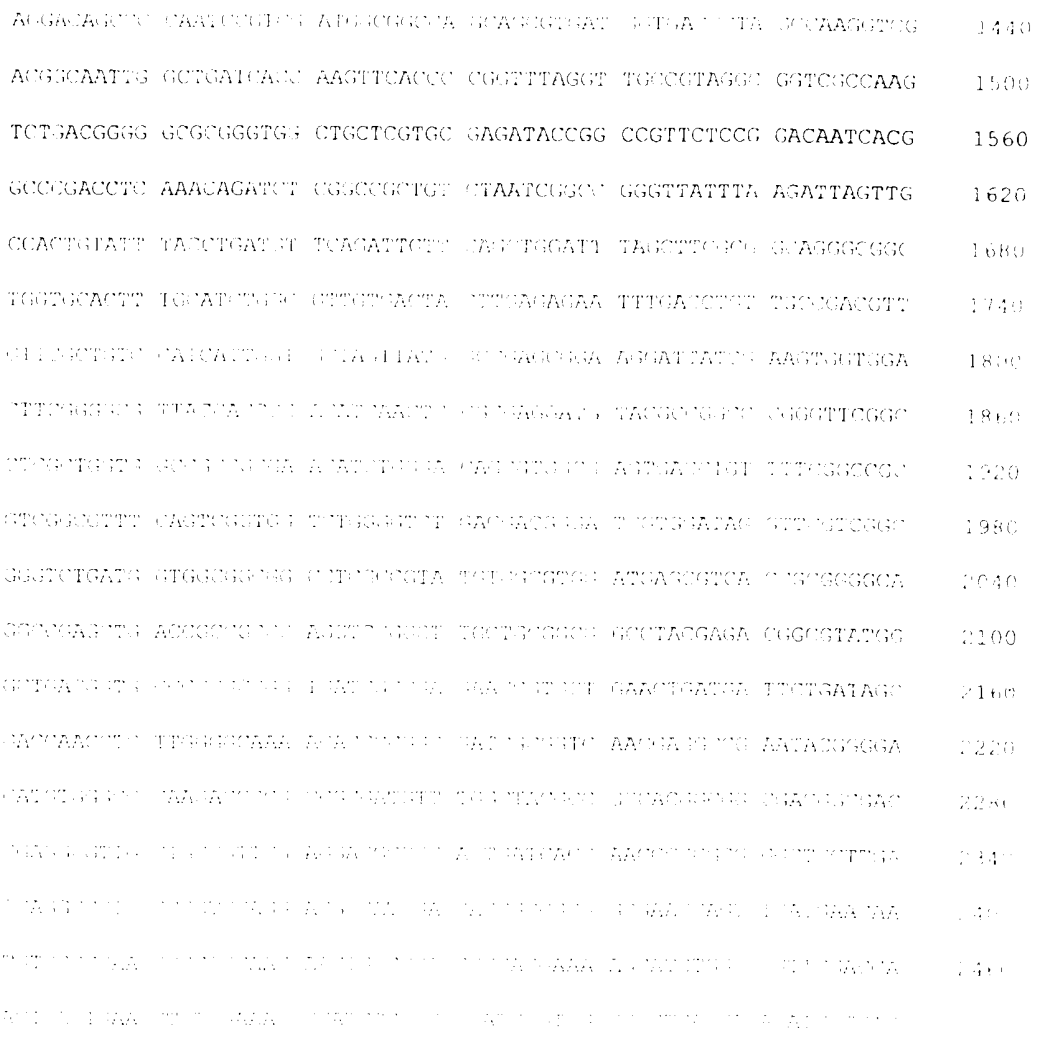
Leu Thr Asn Thr Leu Met Leu Lys Leu Ala Pro Ala
260 265

Ala Ala Gln Ala Val Leu Thr Asn Gly Ile Thr Ala Met
270 275

Ser Ser Ile Leu Ile Leu Gly Leu Leu
280 285

(B) TYPE: [illegible]
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:



(ii) MOLECULE TYPE: Antigen

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:

(2) INFORMATION FOR SEQ ID NO:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:

Val Val Asp Ile Gly Ala Leu Pro Leu Glu Ile Asn Ser Ala Arg Met
1 5 10 15

Tyr Ala Gly Ile Gly Ser Ala Ser Leu Val Ala Ala Ala Tyr Met Trp
20 25 30

Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Glu Ser
35 40 45

Val Val Trp Gly Leu Thr Val Gly Thr Leu Val Ile Gly Val Leu Ala
50 55 60

Leu Met Val Ala Ala Ala Ser Phe Phe Val Ala Trp Met Ser Val Thr
65 70 75 80

Ile Gly Ser Leu Leu Ala Phe Ala Val Leu Gln Ile Phe Ser Asn Ala
85 90 95

Ala Tyr Gln Gln Leu Ala Ala Tyr Ala Ala Ser Arg Glu Ala Ala
100 105 110

Val Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
115 120 125

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
130 135 140

Gln Gln Leu Ala Gln Pro Thr Lys Ser Leu Trp Pro Ile Asp Gln Leu
215 220

Ser Glu Leu Trp Lys Ala Ile Ser Pro His Leu Ser Pro Leu Ser Asn
225 230 235 240

Ile Val Ser Met Leu Asn Asn His Val Ser Met Thr Asn Ser Gly Val
245 250 255

Ser Met Ala Ser Thr Leu His Ser Met Leu Lys Gly Ile Ala Pro Ala
260 265 270

Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Gln Ala Met
275 280 285

Pro Pro Leu Gly Ala Leu Lys Ser Leu Gly Ser Leu Gly Leu
290 295 300

Gly Ala Gly Val Ala Ala Asn Leu Gly Ala Ala Ala Ser Val Lys Ser
305 310 315 320

Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro
325 330 335

Ala Ala Asn Ala Leu Pro Val Thr Asp Leu Thr Ser Ala Gln Thr
340 345 350

Ala Leu Gly His Met Leu Gly Gly Leu Ile Leu Gly Leu Thr Asn
355 360 365

Ser Gly Gly Gly Ile Gly Gly Val Ile Asn Ala Leu Asn Met Pro Pro
370 375 380

Pro Ala Thr Val Met Ile Asp Thr Ile Ala Gly
385 390 395

AGCCTGGACT GGAGTTGCCA CCGACAAGGC GTTCAGGATT GCAAAGCCGA TGGTGGTCTG 240

GCTACAAACC GCGTCAACAC AGCCAAGAC CGTCGCGATG CAGGCGACGG CGCAAGCCGC 300

GGCATACACC CAGGCCATGG CCACGACGCC GTCGCTGCCG GAGATCGCCG CCAACCACAT 360

CACCCAGGCC GTCCTTACGG CCACCAACTT CTTCGGTATC AACACGATCC CGATCGCGTT 420

GACCGAGATG GATTATTTCA TCGTTATGTG GAACTACGCA GCCCTGGCAA TCGAGGTCTA 480

CCAGCCCGAG ACCCTCGTTA ACACGTTTCT CAAGAATTTC GAGCGATGGT CGTCGATCCT 540

TGATCCCGAT GCGAACTCAA TCACTAACAA CGTTATTTTC GAATGCCCT CTTCTGGCAG 600

CTTACTCTCT GTTTCCAACT TCCTCAACTC GTTACTTCAG ACCCTCGGCC AACTGGGTGA 660

GATGAGCGGC CGGATCGAGC AGCTGATCAA CCGCTTCAC CAGGTGACCT CGTTGTTCAG 720

CCAGTTCGGC CGCACCGCGA GCGGCAACTT ATTCGATCAG GAAACTGGCC AGATCAGCCT 780

GCTGGCACC AATCCGCTTT CGAATGATCA GTTCAACTCT GGTCAGGCC CTAGTCCGGG 840

CGTGGGCCTG GTGCCGCCGG AGTCGCTAC TTCTCAGTT AGGTCGTTGA GCCGCACGCC 900

TTTGATCT T AACCTATC AAAAGTCCG TCTGCTGTCG GTCATCGCCG CGAGTTGCTG 960

CGGATCGTCG CGACTGGTG GCGCTTCTC GGACCGATGG GCCAGGGTGC 1020

CAATCCTCG CTTCAACCA TCCTCCGCCC GCCCCCCCC TACCCACTCG CGAGCAGCC 1080

TAATAAAAT AATAAATC ATTTCA ACAAAA AGTTGCATT CTTCAAGTA 1140

AATAAATCT TTTTTA TTTTCA AATTAA ATTTTAA GTTTTAAAA 1200

AAAAAAAT ATTTTTCT TTAAACC AAAAAA TTTTAGTT TTTTAAAC 1260

CTTTTTATT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1320

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

CTAGTGGATG GGACCATGGC CATTTTCTGC AGTCTCACTG CTTTCTGTGT TGACATTTTG 60

GTACGCTGGT GTAAATGAAT CACTAGTTTC AAAACGGTTC TCTCTTCCTA TATGTTTCGG 120

AATTTCTATA CTTTCTTGTG GTCGTAAGAG TTTTCTGTAT TCGGCCCGCA TGACAACCTC 180

TGAAATTTTT TTAAACTTTA TAAACATAAA AAACTTCAAT AGCGACGCAA GTTCAAACTC 240

GTCCATTTCT GTCTTTCCTT ATTCTATTTT AACTTAGCGT TTTTCTATTC GAACATCTCA 300

TGACTTTTAT TTATTTTTA AATGATTTTC AATTTTTTTT TTTCATTTTT TGTTCAGG 360

TTCTTCAGCA CTTCGTTTAT TTCGTATCTT CTCGTTTTCT ATTTTTTTTT GATACCTTGT 420

TATCCTTCCG AA 432

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Thr Gln Ala Met Ala Thr Thr Ile Ser Leu Pro Gln Ile Ala Ala Asn
100 105 110

His Ile Thr Gln Ala Val Leu Thr Ala Thr Asn Phe Phe Gly Ile Asn
115 120 125

Thr Ile Pro Ile Ala Leu Thr Glu Met Asp Tyr Phe Ile Arg Met Trp
130 135 140

Asn Gln Ala Ala Leu Ala Met Glu Val Tyr Gln Ala Gly Thr Ala Val
145 150 155 160

Asn Thr Leu Thr Thr Thr Leu Lys Pro Met Ala Ser Ile Leu Asp Pro
165 170 175

Gly Ala Ser Ile Ile Leu Thr Thr Asn Ile Ile Pro Gly Met Ile Ser Pro
180 185 190

Gly Met Met Thr Pro Val Gly Ile Leu Ile Ala Ala Thr Gln Thr
195 200 205

Leu Gly Gln Leu Gly Gln Met Ser Gly Ile Met Gln Gln Leu Thr Gln
210 215 220

Pro Leu Asn Val Thr Asn Leu Phe Asn Ile Gly Gly Thr Gly
225 230 235 240

Gly Gly Asn Ile Ala Asp Ala Ile Ala Gly Ile Leu Gly
245 250 255

Thr Leu Pro Asn Asp Ile Leu Ala Gly Gly Leu Gly Ile Leu Gly
260 265 270

Ala Gly Ala Ile Leu Asn Asn Gly Gly
275 280

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 700 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ala Glu Met Lys Thr Asp Ala Ala Thr Leu Ala Gln His Ala Gly
1 5 10 15

Asn Ile Glu Arg Ile Ser Ala Asp Ser Lys Thr Gln Ile Asp Gln Val
20 25 30

Leu Ser Thr Ala Gly Ser Leu Asn Lys Trp Arg Gly Ala Ala Gly
35 40 45

Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys
50 55 60

Gln Lys Gln Glu Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly
65 70 75 80

Val Gln Tyr Ser Arg Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser
85 90 95

Gln Met Gly Phe
100

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 700 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

TGACCCGCT AATACGAAAA GAAACGAAT AAAAATGCA CACATACTA CTAAATTTC 300

GCGGTATCG AGCCCGCGAC AAAGCAATC TACGAAATC TCACTTCTAT TATTAATTC 360

CTTGACGAGG GCAAGCAGTC CCTGACCAAG CTCGCA 396

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 88 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

Ile Ser Gly Arg Leu Lys Thr Ser Ile Arg Gln Val Val Ser Thr Ala
1 5 10 15

Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln
20 25 30

Ala Ala Val Val Arg Ile Ala Glu Ala Ala Asp Gln Gln Lys Glu Glu
35 40 45

Leu Arg Glu Ile Ser Thr Asn Ile Arg Ala Ala Gly Val His Tyr Ser
50 55 60

Asn Ala Asp [illegible]
65 70 75 80

(2) INFORMATION FOR SEQ ID NO: [illegible]:

[illegible]

CCTGGTACTC CTCTCAACCT CTACCGCTCC AGCCTCCTCC GAGCTTCTTC AGGGACTGCT 240

TCCCCTCGTC AAGGAGGGAA TGAATGGACG TGACATTTCC CTGGATTGCG CTTGCCGCGG 300

CCTCGATACC CGCGAAATTC CACTGCTGCT CTGTCATGTT TTTGCTCCGT TTCTTTTCGT 360

ATTAGCGGGT CAGAAGCCCA TTTGCGA 387

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

TGGTACGACG ATTCCGTTTA TTTAAACCT CTTCCTAGG GCTTCGTTCC GCGCGTGAGC 60

TGGAGGTCGA ATCTTTCTTC TTCCTCGAC TACCTTGGA TCGATGGAAC ACTTCTTACC 120

TTCCCGACCT TTCTTTCCG CTCTTTCGA TAAGCTCAC CCCGCCGCAC ACTCGCGAG 180

TCTTTTTCT TGGCTCAAC CTCCTTTCG CGAGTTTCT TCATCAACC 240

CATTTTTCT TGCTGACTT ACCTTTTA 268

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

(ii) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
1 5 10 15

2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Ala Ala Met Ser Ala Ala Ile [illegible] Ala [illegible] Ala Ala Leu
1 5 10

2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:
[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gly Lys Ala Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Ala Lys Ala Ser Leu Ser Ile Lys Thr Val Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Pro Lys Thr Tyr Met Gln Glu Leu Lys Gly Thr Asp Thr Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser
1 5 10 15

Leu Leu Asn Ala Asn Leu Ala Ala Pro Ser Val Ser Leu Ala Asn

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(C) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Asp Pro Gly Tyr Thr Pro Gly
1 5

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) OTHER INFORMATION: /note= "The Asp at position 1 can be either a D-Asp or Thr"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Xaa Xaa Gly Tyr Thr Pro Gly
1 5

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) OTHER INFORMATION: /note= "The Asp at position 1 can be either a D-Asp or Thr"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Xaa Xaa Gly Tyr Thr Pro Gly
1 5

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Xaa Xaa Xaa His Lys Pro Phe Leu Arg

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Xaa Asp Ser His Lys Xaa Ala Phe Xaa Xaa Xaa Xaa Xaa Xaa Xaa

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Ala Pro Glu Ser Gly Ala Gly Leu Gly Gly Thr Val Gln Ala Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Xaa Tyr Ile Ala Tyr Xaa Thr Thr Ala Gly Ile Val Xaa Gly Lys Ile
1 5 10 15

Asn Val His Leu Val
20

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 957 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS:
(D) TOPOLOGY:

GCCCGGTAC CTCACCGCCG GTGGTCGCGA CCCCGGCTTC GCCTGCGCCG GAAGCGCCGG 300

CCGTGCCCGG TGTTGTGCCT GCCCCGGTGC CAATCCCGGT CCCGATCATC ATTCCCCCGT 360

TCCCGGGTTG GCAGCCTGGA ATGCCGACCA TCCCCACCGC ACCGCCGACG ACGCCGGTGA 420

CCACGTCGGC GACGACGCCG CCGACCACCC CGCCGACCAC GCCGGTGACC ACGCCGCCAA 480

CGACGCCGCC GACCACGCCG GTGACCACGC CGCCAACGAC GCCGCCGACT ACGCCGGTGA 540

CCACGCCACC AACGACCGTC GCCCCGACGA CCCTCGCCCC GACGACGTTC TTCCCGACCA 600

CCTTCGCCAC GACGACGTTC GCTCCAGCCA CGCCGACGCT GACGACCGTC GCTCCGCAGC 660

CGCCGTACTC GCCCATCTAA TAACCAATTT AAAAGAGTTG TAACCTAGTA TATATGTGG 720

CCCCGCAGAC GGTGGCGCCG GCTTCGCAGT CGCATTCTTT TGCCCGCAAC ACCAGCGGCG 780

CCGCCGACTT ATTTGCCGAG TTCCGATGAT CTTCGCCGTT TCATTATCGT TAAGGGACAT 840

GCCCCGTTGAT GCGGTGACGG TGGTGCTGCT CTGTTTCTAAC GA 882

( ) INFORMATION FOR SEQ ID NO:144:

( ) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 810 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

( ) MOLECULE TYPE: DNA (genomic)

( ) SEQUENCE DESCRIPTION: SEQ ID NO:144:

AGCGGCTAGC TGTCGGGTTT CGTCAAGCTT GATCATGATG TCGAGGTGAC TTTGACGCCG 600

CCCCCGAAG GAGGCGCTGA ACTCGGCGTT GAGCCGATCG GCGATCGGTT GGGGCAGTGC 660

CCAGGCCAAT ACGGGGATAC CGGGTGTCNA AGCCGCCCCG AGCGCAGCTT CGGTTGCGCG 720

ACNGTGGTCG GGGTGGCCTG TTACGCCGTT GTCNTCGAAC ACGAGTAGCA GGTCTGCTCC 780

GGCGAGGGCA TCGACCACGC GTTGGCTCAC CTCGT 815

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1152 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

ACAAGGCGCT GGTTGAGGTC TCAGATCAAA GATCTTCTGA ACTCACGGCA GGGTTCAAGC 60

CTTCTCCCAG AACAACTGCT GAAGATCTAC CGGCCGAAA CAGCGCTGA TTTAGCGTC 120

TATGACAAGT TCAATTACTA GATCATCGCA TAATTATA TGGCAGCGCT GGCTAACAA 180

TTTTAAAAA TTTTTAATC TATGCTTA AATCAGCAT TTTGATGCA TTTGCGGGA 240

ATTTTTTA TTTTTTAA TGATAATC TTTAAAGT ATGTTCAA GAAATTTA 300

ATATTTTA TTTTTTTT TGTTCAGA ATAAATGT TTCTAATC TAAATGTC 360

GCTGAAACCG CTGGCTGGCT CTGAGTGTCT GTACGTAAT CCGTTGCGAA CAGGCCGGCC 960

CGCCGGCCGA ATACCAGCAG ATCGGACAGC GAATTGCCGC CCAGCCGGTT GGAGCCGTGC 1020

ATACCGCCGG CACACTCACC GGCAGCGAAC AGGCCTGGCA CCGTGGCGGC GCCGGTGTCC 1080

GCGTCTACTT CGACACCGCC CATCACGTAG TGACACGTCC GCCCGACTTC CATTGCCTGC 1140

GTTCGGCACG AG 1152

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 655 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

CTTCTCCGGA TTGGCAGGG TGTACTTCTT GCTTCTTAA GCGCCATAAC TCTCGAACGAC 60

CAGCAATCGG GCAACAGCA AGATCCTCTT GATCATCTT ACACCGTTCA CTCGGCGCAT 120

CGCTTCGAGT CCTCCTCCG CGCTTTTTC TTCTTCAAG TTACCAAAC CTTTTCTCGC 180

CGTCTTTAG CTCTCATCAT CCCTTCTCG TCTTCTTCA TTCATCATCA TTGAGCTTCA 240

ACTTCAACA CCAACAACA TCTTCTCCA ACTTTTCC ACCGCCTT TGCACCTT 300

CTTCAAGT CCATTACCA GAACTCTGTC GAAGCCTTA ACTACCAAG TCTTCTCGCC 360

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Asn Ala Val Val Ala Phe Ala Val Ile Gly Phe Ala Ser Leu Ala Val
1 5 10 15

Ala Val Ala Val Thr Ile Arg Pro Thr Ala Ala Ser Lys Pro Val Glu
20 25 30

Gly His Gln Asn Ala Gln Pro Gly Lys Phe Met Pro Leu Leu Pro Thr
35 40 45

Gln Gln Gln Ala Pro Val Pro Pro Pro Pro Pro Pro Asp Asp Pro Thr Ala
50 55 60

Gly Phe Gln Gly Gly Thr Ile Pro Ala Val Gln Asn Val Val Pro Arg
65 70 75 80

Pro Gly Thr Ser Pro Gly Val Gly Gly Thr Pro Ala Ser Pro Ala Pro
85 90 95

Val Ala Pro Ala Val Ile Gly Val Val Ala Ala Pro Val Pro Ile Pro
100 105 110

Val Pro Phe Ile Ile Ile Ile Pro Pro Pro Ile Gly Trp Gln Pro Gly Met Pro
115 120 125

Phe Ile Pro Thr Ala Pro Ile Ile Ile Val Thr Ile Ser Ala Thr
130 135 140

Val Pro Leu Thr Ile Pro Ile Ile Val Gln Thr Val Ile Ile Pro Thr
145 150 155 160

Ile Ile Ile Ile Ile Ile Ala Ile
165

Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile
170 175 180

Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile
185 190 195 200

Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile Ile
205 210 215 220

Gly Ser Gly Gly Gly Asp Leu Phe Gly Gly Phe
260 265

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

Ile Asn Glu Pro Leu Ala Pro Pro Ala Pro Pro Asp Pro Pro Ser Pro
1 5 10 15

Pro Arg Pro Pro Val Pro Pro Val Pro Pro Leu Pro Pro Ser Pro Pro
20 25 30

Ser Pro Pro Thr Gly Trp Val Pro Ala Ala Leu Leu Pro Pro Trp Leu
35 40 45

Ala Gly Thr Pro Pro Ala Pro Pro Val Pro Pro Met Ala Pro Leu Pro
50 55 60

Pro Ala Ala Pro Leu Pro Pro Leu Pro Pro Leu Pro Pro Leu Pro Thr
65 70 75 80

Ser His Pro Pro Arg Pro Pro Ala Pro Arg Pro Pro Ala Pro Leu Pro
85 90 95

Ala Ser Pro Pro Val Pro Val Ala Ala Pro Pro Pro Ala Ala Ala Ala
100 105 110

Pro Pro Pro Pro Ala Ala Ala Pro Pro Ala Ala Ala Ala Pro Pro Ala
115 120 125

Arg Ala Ala Ile Gly Ala Leu Ala Leu Pro Leu Ala Gly Leu Leu
130 135 140

(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Xaa Pro Pro Ala Glu Val Ser Asp Thr Arg Val Ser Gly Leu Thr Gly
1 5 10 15

Ala Val Gln Pro Ser Pro Arg Thr Thr Ala Glu Asp Pro Arg Pro Arg
20 25 30

Asn Arg Arg
35

(2) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

103

(2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

GAATCCATAT GGGCCATCAT CATCATCATC ACGCGATCGA CATCATCTCG ACC 53

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

GGATCCTGCA GGCTCGAAAC CACCGAGCGG T 31

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

CTCTGAATTC ACCGTTGAA ATTTCGAGA T 31

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(iii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(v) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

GAGAGAATTC TCAGAAGCCC ATTTGGGAGT A A 33

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1950 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(v) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1950

(x) PUBLICATION INFORMATION: (J. H. Rothman

```
TTT CAC GAC AGG TAT CCG AAC GTC ACG ATC ACC GCT CAG GGC ACC GGT     412
Phe His Glu Arg Tyr Pro Asn Val Thr Ile Thr Ala Gln Gly Thr Gly
             75                  80                  85

TCT GGT GCC GGG ATC GCG CAG GCC GCC GCC GGG ACG GTC AAC ATT GGG     460
Ser Gly Ala Gly Ile Ala Gln Ala Ala Ala Gly Thr Val Asn Ile Gly
         90                  95                 100

GCC TCC GAC GCC TAT CTG TCG GAA GGT GAT ATG GCC GCG CAC AAG GGC     508
Ala Ser Asp Ala Tyr Leu Ser Gln Gly Asp Met Ala Ala His Lys Gly
    105                 110                 115

CTG ATG AAC ATC GCG CTA GCC ATC TCG GCG CAG CAG GTC AAC TAC AAC     556
Leu Met Asn Ile Ala Leu Ala Ile Ser Ala Gln Gln Val Asn Tyr Asn
120                 125                 130                 135

CTG CCC GGA GTG AGC CAG CAC CTC AAC CTG AAC GGA AAA GTC CTG GCC     604
Leu Pro Gly Val Ser Gln His Leu Tyr Leu Asn Gly Tyr Val Leu Ala
                140                 145                 150

GCC ATG TAC CAG GGC ACC ATC AAA ACC TGG GAC GAC CCG CAG ATC GCT     652
Ala Met Tyr Gln Gly Thr Ile Lys Thr Trp Asp Asp Pro Gln Ile Ala
            155                 160                 165

GCG CTG AAC CTC GGC GTG AAC CTG CCC GGC ACC GCG GTA GTT CCG CTG     700
Ala Leu Asn Ile Gly Val Asn Leu Pro Gly Thr Ala Val Val Pro Leu
        170                 175                 180

CAC GAC TCC GAC GGC TCC GGC GAC ACC ATC CTG TTC ACC CAG TAC CTG     748
His Asp Ser Asp Gly Ser Gly Asp Thr Ile Leu Phe Thr Gln Tyr Leu
    185                 190                 195

[illegible]
Ser Tyr Gln Asn [illegible] Gly Thr
```

Gln Ala Ile Ser Met Ile Asp Gly Pro Ala Pro Asp Gly Tyr Pro Ile 300 305 310	1084
ATC AAC TAC GAG TAC GCC ATC GTC AAC AAC CGG CAA AAG GAC GCC GCC Ile Asn Tyr Glu Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala 315 320 325	1132
ACC GTG CAG ATC TTC CAG GCA TTT [illegible] CAC TGG Trp Ala Ile 340	1180
AAC AAG GCT TCT TTC [illegible] GAC GTT CAC TTC CAG CCG CTG CCG CCG Asn Lys Ala Ser [illegible] Asp Val His Phe Gln Pro Leu Pro Pro 345 350 355	1228
GTC GCC GTG AAG [illegible] GAT GCA [illegible] ATC [illegible] ATT [illegible] AG Val Ala Val Lys [illegible] Asp Ala [illegible] Ile [illegible] Ile [illegible] Ser Ser 360 365 370	1276
TAGCTTGTTT GAGAAGGATA GCATAGTAAA [illegible] [illegible] TGGTCT GTTTGTGGCA	1333
GTATCTTGGC [illegible] [illegible] [illegible] [illegible] [illegible]	1384
TATAAGGGT [illegible] [illegible] [illegible] [illegible] [illegible]	1435
ATGTGATTTT TTGTATGAAA [illegible] [illegible] [illegible] [illegible]	1486
AGGAAAAA [illegible] [illegible] [illegible] [illegible] [illegible]	1537
TACTGGCTT [illegible] [illegible] [illegible] [illegible] [illegible]	1588
[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]	1639
[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]	1690

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

Val Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
1 5 10 15

Leu Leu Leu Ala Ala Ala Gly Cys Gly Ser Lys Pro Pro Ser Gly Ser
20 25 30

Pro Glu Thr Gly Ala Glu Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
35 40 45

Pro [illegible] Val Thr Leu Ala [illegible] Thr [illegible] Leu Leu Tyr Ile Leu
50 55 60 65

Leu Ala Leu Tyr Gly Pro Ala Thr His Leu Arg Tyr Pro Asn Val Thr
70 75 80

Ile Thr Ala [illegible] Gly Thr [illegible] Ser [illegible] Leu Ala Gln Ala Ala
85 90 95

Ala Gly Thr Val Asn Ile Gly Ala Ser Asn Ala Tyr Leu Ser Gln Gly
100 105 110

Asp Met Ala Ala Leu Lys Gly Leu Met Asn Leu Ala Leu Asn Ile Ser
115 120 125

Asn Ile Ile Val Asn Tyr Asn Leu Leu Asp Val Phe Thr Phe Leu Lys
130 135 140

Leu Asn Lys Lys Val Leu Ala Asn Met Leu Thr Gly Thr Tyr Asn Thr
145 150 155 160

Thr Asn Pro Ile Ile Ala Leu Leu Ser Ile Phe Pro
165 170

Phe Leu Leu Pro Asn Asn Gln Ser Leu Gln Ala Ala Asn Ala Ala Gly Phe
275 280 285

Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
290 295 300

Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
305 310 315 320

Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
325 330 335

Arg Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
340 345 350

Lys Phe Gln Pro Leu Pro Pro Asn Val Val Lys Leu Ser Asp Ala Leu
355 360 365

Leu Ala Thr Ile Ser Ser
370

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1393 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible] 660

[illegible] 720

TGACACCTTC TTGTTCACCC AGTACGTGTC CAAGCAAGAT CCCGAGGGCT GGGGCAAGTC 780

GCCCGGCTTC GGCACCACCG TCGACTTCCC GGCGGTGCCG GGTGCGCTGG GTGAGAACGG 840

CAACGGCGGC ATGGTGACCG GTTGCGCCGA GACACCGGGC TGCGTGGCCT ATATCGGCAT 900

[illegible] 960

[illegible] 1020

[illegible] 1080

[illegible] 1140

[illegible] 1200

[illegible] 1260

[illegible] 1320

[illegible] 1380

[illegible] 1440

[illegible] 1500

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

Met Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
1 5 10 15

Leu Leu Leu Ala Ala Ala Gly Gly Gly Ser Lys Pro Pro Ser Gly Ser
20 25 30

Pro Gln Thr Gly Ala Gly Ala Gly Ser Val Ala Thr Thr Pro Ala Ser
35 40 45

Ser Pro Val Thr Leu Ala Gln Thr Gly Ser Thr Leu Leu Tyr Pro Leu
50 55 60

Pro Asn Leu Lys Lys Ile Ala Ile His Ser Arg Tyr Ile Asn Val Thr
65 70 75 80

Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Leu Ala Gln Ala Ala
85 90 95

Ala Gly Thr Val Asn Leu Gly Ala Ser Asp Ala Tyr Leu Ser Gln Gly
100 105 110

Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser
115 120 125

Gly Gln Val Asn Tyr Arg Leu Ile Val Thr Gln His Ser Lys
130 135 140

Val Asn Leu Lys Thr Leu Ala Ala Ser Lys Gln Ser Thr Tyr Pro
145 150 155

Thr Asn

Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Pro Gly Asn
260 265 270

Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
275 280 285

Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
290 295 300

Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Gln Tyr Ala Ile Val Asn
305 310 315 320

Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
325 330 335

Arg Ala Ile Thr Arg Gly Asn Lys Ala Ser Phe Leu Asp Glu Val
340 345 350

His Ala Gln Pro Leu Pro Pro Arg Val Val Lys Leu Ser Asp Ala Leu
355 360 365

Ile Ala Thr Ile Ser Ser
370

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1777 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

TGCGATCTGA TCGGGATGGG CGCGTCGGAC AAGCTCAGCC CATCGGGACC CGACGGCTAT 660

AGCTATGGCG AGCAACGAGA CTTTTTGTTC GCGCTCTGGG ATGCGCTCGA CCTCGGCGAC 720

CACGTGGTAC TGGTGCTGCA CGACTGGGGC TCGGCGCTCG GCTTCGACTG GGCTAACCAG 780

CATCGCGACC GAGTGCAGGG GATCGCGTTC ATGGAATCGA TGGTCACCCC GATGACGTGG 840

[illegible] 900

[illegible] 960

[illegible] 1020

[illegible] 1080

[illegible] 1140

[illegible] 1200

[illegible] 1260

[illegible] 1320

[illegible] 1380

[illegible] 1440

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1574:

GAGATTGAAT CGTACCGCTC TCCTTAGCGC CTCCTTCCCC TGAATGCCCA TATCACGCAC 60

GCCCATGTTC TGGCTGTCGA CCTTCGCCCC ATGCCCGGAC GTTGGTAAAC CCAGGGTTTG 120

ATCAGTAATT CCGGGGGACG GTTGCGGGAA GCCGGCCAGG ATGTGCGTGA GCCGCGGCGC 180

CGCCGTCGCC CAGGCCACCG CTGGATGCTC ATCCTCGGTG CGGCGACGTA GCCAGGCGTT T 241

GGTCGTTGTC GTCTACATCG GTACTCGGTT GATACGCAT CGGCGTTCAT CTCTCGAAGAC 300

CCTGATCGAC GCCCCGATT CAGA 324

(2) INFORMATION FOR SEQ ID NO:1575:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1348 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1575:

CCGACATAC CCACGAGTCA GCCTACCGC CTCCGCACC GAAAGTTGTT CCTCCACTGC 900

CGGTTAACCG GCCAGAGGGT CTGGACGCCA TCGAGGATCT GCACATCTGG ACCGCCGAGT 960

CGGTGCGCGC CGACCGGCTC GACTTTCGGC CCAAGCACAA ACTGGCCGTC TTGGTGGTCT 1020

CGGCGATCCC GCTGGCCGAC CCGGTCCGGC TGGCGCGTAG GCCCGAGTAC GGCGGTTGCA 1080

CCAGCTGGGT GCACTTCCTC CTGACGCCGA CGTTTGCGG GCCGGTTGAC GACGAGGCCG 1140

CGCTCCTCGA CGTTCTCCTC GCCTTCTCCA AGTCCTTCTC CTGACTTCCT CCTATCGCTT 1200

CCTTTTAACC TCTACGCTTA CTCCCCTCC CGACCATCT GTCATCCTT CCTTCCTGC 1260

CCCCTCAAT TGATCCCTC CCAACACCA CATTGGTGT CGTCATCTTC TTCCTCGCCG 1320

CCCCCACTC TACAAC 1338

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 521 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

GAAGACCCGG CCCCGCCATA TCGATCGGCT CGCCGACTAC TTTCGTCGAA CGTGCACGCG 60

GCGGCGTCGG GCTGATCATC ACCGGTGGCT ACGCGCCCAA CCGCACCGGA TGGCTGCTGC 120

CGTTCGCCTC CGAACTCGTC ACTTTCGGTG AAGCCCGACG GCACCGCCGA ATCACCAGGG 180

CGGTCCACGA TTCGTTTTCA AAGATCCTGT TCGAAATCGT GCACGCCGGA CGCTACGCCT 240

ACTACTACT TCGGTTACT TTTTCGTGGA TGAAGTCGTC GATCACCTTT TTCGTCCGT 300

GAGATATTT CGGTCGAGT TTGAAGTCGA CATCTTCGA TTTCGTCCGT TGCCCGCAGT 360

TCTTTCGGA TGCTTCGTA TATTCGTCG AAATCATCGT CATCGAAGTC TATCTGCTCA 420

AAACTTATT ACCTCGCTT ACTAATAAAT GTATCATTT TTGCATCTT ACACCTGCCA 480

ACGGTCGCCT GT 492

(2) INFORMATION FOR SEQ ID NO:156:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 98 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

Gly Asp Ala Ile Val Phe Gln His Gly Asn Pro Thr Ser Ser Tyr Leu
115 120 125

Trp Arg Asn Ile Met Pro His Leu Glu Gly Leu Gly Arg Leu Val Ala
130 135 140

Cys Asp Leu Ile Gly Met Gly Ala Ser Asp Lys Leu Ser Pro Ser Gly
145 150 155 160

Pro Asp Arg Tyr Ser Tyr Gly Gln Gln Arg Asp Phe Leu Phe Ala Leu
165 170 175

Trp Asp Ala Leu Asp Leu Gly Asp His Val Val Leu Val Leu His Asp
180 185 190

Trp Gly Ser Ala Leu Gly Phe Asp Trp Ala Asn Gln His Arg Asp Arg
195 200 205

Val Gln Gly Ile Ala Ile Met Glu Ala Ile Val Thr Pro Met Thr Trp
210 215 220

Ala Asp Trp Pro Pro Ala Val Arg Gly Val Phe Gln Gly Phe Arg Ser
225 230 235 240

Pro Gln Gly Glu Pro Met Ala Leu Glu His Asn Ile Phe Val Glu Arg
245 250 255

Val Leu Pro Gly Ala Val Leu Arg Gln Leu Ser Asp Glu Val Met Asn
260 265 270

His Tyr Arg Glu Pro Phe Val Asn Pro Val Asp Arg Arg Pro Leu Trp
275 280 285

Arg Phe Pro Asn Glu Leu Pro Ile Ala Gly Glu Pro Ala Asn Ile Val
290 295 300

Leu Arg Leu Pro Ala Pro Gly Arg Asp Leu Gln Gly Lys Gly His Gln
405 410 415

Ser Gln Pro Leu Pro Ser Gln Arg Gly Arg Gln Ile Tyr Val Ala Gly
420 425 430

Gln Arg Ser Ser Tyr Leu Pro Ser Glu Leu Val Ala Ala Phe Leu Trp
435 440 445

Ala Gln Thr Glu Glu Ala Gln Arg Ile Thr Arg Leu Arg Leu Asp Leu
450 455 460

Trp Asn Arg Tyr His Gln Ser Phe Gln Ser Leu Gln Gln Arg Gly Leu
465 470 475 480

Leu Arg Arg Pro Ile Ile Pro Thr Gly Cys Ser His Asn Ala His Met
485 490 495

Tyr Tyr Val Leu Leu Ala Pro Ser Ala Arg Asn Gln Gln Val Leu Ala
500 505 510

Arg Leu Thr Ser Gln Gly Ile Gly Ala Val Phe His Tyr Val Pro Leu
515 520 525

His Asp Ser Pro Ala Gly Arg Arg
530 535

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 541 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

65 70 75 80

Gly Gly Leu Thr Val Asp Trp Lys Val Ser Trp Pro Arg Gln Arg Gly
85 90 95

Ala Thr Val Leu Ala Ala Val His Glu Trp Pro Pro Ile Val Val His
100 105 110

Phe Leu Val Ala Glu Leu Ser Gln Asp Arg Pro Gly Gln His Pro Phe
115 120 125

Asp Leu Asp Val Val Leu Gln Arg His Trp Leu Ala Leu Arg Arg Ser
130 135 140

Glu Thr Leu Glu His Thr Pro His Gly Arg Arg Pro Val Arg Pro Arg
145 150 155 160

His Arg Gly Asp Arg Ile Ala His Arg Asn Pro Leu His Ser Val
165 170 175

Ala Met Leu Val Ser Pro Val Leu Ala Glu Arg Arg Ala Pro Val Val
180 185 190

Glu His Gln Tyr His Val Val Ala Thr Val Gln Arg Ile Pro Glu Arg
195 200 205

Glu Gln Leu Val Ser Leu Leu Arg Ile Ala Ile Ala Val Gly Ser Arg
210 215 220

Trp Ala Glu Leu Val Arg Ala Ala His Pro Asp Glu Ile Arg Gly His
225 230 235 240

ATGAACATCT CGTCGGTCGT CGGTTGCAAG CTTTTGTTAC GATTCGCCGT TTACTCTTCC 60

GCCATGCACG CGATCGCCGG TTTCTCCGAT GCGTTGGGAC AAGAGCTGCT CGGTAGCGGA 120

ATCGCCGTCT CGGTGATCCA CCCGGCGCTG ACCCAGACAC CGCTGTTGGC CAACGTCGAC 180

CCCGCCGACA TGCCGCCGCC GTTTCGCAGC CTCACGCCCA TTCCCGTTCA CTGGGTCGCG 240

GCAGCGGTGC TTGACGGTGT GGCG 264

(2) INFORMATION FOR SEQ ID NO:1594:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1594:

[illegible]

GAGTACACCA ACATCGGCGC CTCCATCGTG TTCCTGAACT ACGACGTCAT CTTCAAAACG 1020

CGCACCACCG TCGGTTTCGA CGTACTGGAC GGGTCCGACA CCATGTTCGT GGCCCCAGTA 1080

ACCATCGGCG ACGGCGCGTA TACCGGGGCC GGCACAGTGG TCCGGGAGGA TGTCCCGCCG 1140

GGGGCGCTGG CAGTGTCGGC GGGTCCGCAA C 1171

(2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:160:

TAAAGCGCCA CACGCTGAGC GCCCCATCCA ACGCTTCAAA GCCAGTTGCA GCCACCCAAG 60

ACTCTGGCCA AGGCTCCATC GGCCTATCCT CCTCGAATCT CTGCCCCGCT CGTACCTACT 120

TCACTCAACG TCGCACTCCT AAGCCTCCCA ACTCCTCTGA CCCCGAGCTT CTCCTCGAAC 180

[illegible] 240

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) [illegible]
(B) [illegible]
(C) [illegible]
(D) [illegible]

CGCC 404

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1439 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

[illegible] 1260

GCGGCCAGGG CGGTGCCGGC GGCAGCGCCG GCAACGGCGG CCACGGCGGC GGTGCCACCG 1320

GCGGCGCCAG CGGCAAGGGC GGCAACGGCA CCAGCGGTGC CGCCAGCGGC TCAGGCGTCA 1380

TCAACGTCAC CGCCGGCCAC GGCGGCAACG GCGGCAATGG CGGCAACGGC GGCAACGGC 1439

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible]

[illegible]

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

[illegible]

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

GGGCTGTGTT GCACTCACCA [illegible] TTAACGTTG TCCCCCAAT ATTCAGCTCA 60

AGTCGTACTA TTTACGTTCG CAGAACCGCC GTATCAAGGT GCGGGTCAGC GCCCAAGGAA 120

TCAAGGTCAT TTATCGCGAC GTTCTTCTAT [illegible] CAGCGTCCTT CAGGATCCC 180

[illegible] 240

CACAACTCAT GCCCACCAAC GAGATTCGTT CAATTCTCAA [illegible] AACCAAGGGA 300

CCCTTATACT GTATCTTAAT CTATCTCACT TTCAAGAATC TTAATAACTA TTCTTGTATC 360

CCCTTACAGT TTAAGAGTTC AAAACTCTTA TA 392

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible] 540

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible]

[illegible] 60

[illegible] 120

[illegible] 180

GACGAGCTAC GCCGATGTGC GAACGCCGTT [illegible] 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 660

[illegible] 60

TGGGCACAGC CGGGGTCGGC GGCAACGGTG GCGCCGCAGC TACCGTTATT GCAGTTGGCG 120

GCTCTGGGGG CACCGGCGGC GACGGCGGGA CCGGCGGGCG TGGCGGCGTG TTAATGGGCG 180

CCGGCGCCGG CGGGCACGGT GGCACTGGCG GCGCGGGCGG TGCCGGTGTC GACGGTGGCG 240

GCGCTGGCGG GGCCGGCGGG CCCGGCGGCA ACTGCGGCGC CGGGGGTCAA GCCGCCCTGC 300

TGTTCGGCGC CGCCTTCACC GCCGAGCCC GCCTCTACGG TGGGGATGCC GGTGCCGGCG 360

GTGATCGTTT CGATCTTAAC ATTGCGGCT CCCTCGCCAC CGGCGC 406

(2) INFORMATION FOR SEQ ID NO:1-4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1-4:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

GGTGGTAACG GCGGGCAGGG TGGCATCGGC GGCGCCGGCG AGAGAGGCGC CGACGGCGCC 60

GGCGCCAATG CTAACGGCGC AAACGGCGAG AACGGCGGTA GCGGTGGTAA CGGTGGCGAC 120

GGCGGCGCCG GCGGCAATGG CGGCGCGGGC GGCAACGGCG AGGCGGCCGG GTACACCGAC 180

GGCGCCACGG GCACCGGCGG CGACGGCGGC AACGGCGGT 219

(2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 494 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

GGGCCGGTGG TGCCGGGCCG CAGTGTTTCA GCGTGGACGG CGAGGCGGCT TCGTTGGGGG 60

TTGGCGGCAC CGCCGGCCAG GGTGGGGCTG GCGGTGCCGG AGCGGCCGGC GCCGACGCCC 120

CCGCCAGCAC AGGTCTAACC GGTGGTACCG GGTTCGCTGG CGGGGCCGGC GGCGTCGGCG 180

GCCAGAGCGG CAACGCCATT GCCGGCGGCA TCAACGGCTC 220

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

[illegible] 360

[illegible] 420

TCACCTAAAA CTCTGACCGC AACCCCCCCA ACCCCCCCA CCCCCCCCCC CCCCCAACCC 19

GCGAAAACGG CGGAAACGGC GCAGACAACA CCACCACCGC CGCCGCCGGC ACCACAGGCG 240

GCGACGGCGG GGCCGGCGGG GCCGGCGGAA CCGGCGGAAC CGGCGGAGCC GCCGGCACCG 300

GCACCGGCGG CCAACAAGGC AACGGCGGCA ACGGCGGCAC CGGCGGCAAA GGCGGCACCG 360

GCGGCGACGG CAACTCTCCA GGCAGCACCC CCACCGCGCC 400

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 536 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

CCAACCCC CCAACCCC CC CACCCC CCC CCCC CCC AACCCCCCCC CCCACCCC 60

ACCCCCAACC CCCCCCAACC CCCCCCCCCC CCCCCAACCC CCCCAACCCC CCATCCCCCC 120

CAACCCCCC CCCCCCCACC CCCACCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCAC 180

CCC CC C CCCCCCAACC CCCCCCCCCA CCCCCCCCCC CCCCCCCCCC CACCCCCCCC 240

CCCCCCCCCC CCCCCCCCCC CCCAACCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC 300

CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC 360

CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC 420

CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC 480

CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCC 536

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

GAGACCGTGG TGCCCCCGGC CAGTTCTTCA GCGCCGAGGG TGCGGCTCGT GCGTTTCGCG 60

TTGGCGGCAC CGGCGGCCAG GGTGGGGCTG GCGGTGCCGG AGCGGCCGGC GCCGACGCCC 120

CCGCCAGCAC AGGTCTAACC GGTGGTACCG GGTTCGCTGG CGGGGCCGGC GGCGTCGGCG 180

GCCACGGCGG CAACGCCATT GCCGGCGGCA TCAACGGCTC CGGTGGTGTC GACGGCACC 239

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

AGCAGCGCTA CTGGTGGTGG CGGGTTCTAT GTGGAAAACG GCGAGCTGGC 60

GGCAACAGGC GTGTGGTGGC TACTAAAGTC GATCTGACGG CTAATGGCAA G 111

CGTACTGCCC CGAACACTTA GAAGA 945

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2138 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

CTTTCCAATT ATTCGTATTT TTTTTATTT TAATTCTAAA ATTCACTTTT TTTCTTTACT 60

GGAGCAAAAT CCTATCAAAA TGGCTTTT GAAGCATTGT GATCATTTTT CTATCCTTGC 120

ATGAACGGGC GCCATCAAAT TATTCAAGCA ACTTTTACT TTACTTACTA TAATCGCTAT 180

ACCACTAACG ACCATGATCC GATATGACTT ATTTCAAGAC CTCAACGTTG GATCAGCAAG 240

AGATTTTGAA CAAGCTCAAA GAGCTCAGCA CAAGCATCAT ATTGATCTCC 300

TTATCACATT GTCCAACTT ACTCCCGTTA AAAACCTCCT TAACACGTTT TATTTTTCCG 360

CCAAAAAAT CAATGAATA TTTATCAGTC TCAAAAAGA CTGCATTTT TAATTAACTT 420

TCTTCTTAA TTTTTTTAT TTAATTTCA ATTTTTCAA TTTTTATTTC ATTTTTCTCC 480

ACAACTATTT TAATAACTT CTTATTTTA ACTTTTTTTA TTTTTTTTA TTTTTATTT 540

TAT TAAT AATTCATA TTTTTTTTT TTTATTTTT TTAATTTAA TTTATTAAT 600

TTAAAAAAT TTTTAAA TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT 660

[illegible]

35 40 45

Leu Val Leu Ser Ala Asp Asn Met Arg Gln Tyr Leu Ala Ala Gly Ala
50 55 60

Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Ala
65 70 75 80

Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly
85 90 95

Glu Gly Thr Val Gln Ala Gln Ser Ala Gly Ala Val Gly Gly Asp Ser
100 105 110

Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro
115 120 125

Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp
130 135 140

Gln Gly Ala Ser Leu Ala His Phe Ala Asp Gly Trp Asn Thr Phe Asn
145 150 155 160

Leu Thr Leu Gln Gly Asp Val [illegible] Arg Phe Arg Gly Phe Asp Asn Trp
165 170 175

[illegible] Gly Asp Ala Ala Thr Asn Tyr [illegible] Asn [illegible] Leu Asn Gln Glu Arg
180 185 190

Gln Trp [illegible] Met [illegible] Ala Lys Gln
195 200 205

[illegible] Tyr Val Ala Gln Leu [illegible] Val [illegible] Asn [illegible] His [illegible]

[illegible]
240

[illegible]

Ala Gln Leu Thr Ser Ala Gly Arg Glu Ala Ala Ala Leu Ser Gly Asp
340 345 350

Val Ala Val Lys Ala Ala Ser Leu Gly Gly Gly Gly Gly Gly Gly Val
355 360 365

Pro Ser Ala Pro Leu Gly Ser Ala Ile Gly Gly Ala Glu Ser Val Arg
370 375 380

Pro Ala Gly Ala Gly Asp Ile Ala Gly Leu Gly Glu Gly Arg Ala Gly
385 390 395 400

Gly Gly Ala Ala Leu Gly Gly Gly Gly Met Lys Met Pro Met Gly Ala
405 410 415

Ala His Gln Gly Gln Gly Glu Ala Lys Ser Lys Gly Ser Gln Gln Glu
420 425 430

Asp Glu Ala Leu Lys Trp Arg Asp Ala Ala Trp Thr Ala Arg Val Ile
435 440 445

Gly Asn Asn Ala Arg Arg Asp Ser Lys Ala Ser Lys
450 455 460

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 277 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

Arg Asp Glu Ser Leu Leu Leu Arg Arg Arg Gly Arg Val Asp Leu Asp
100 105 110

Gly Gly Gly Arg Leu Arg Arg Val Tyr Arg Phe Gln Gly Cys Leu Val
115 120 125

Val Val Phe Gly Gln His Leu Leu Arg Pro Leu Leu [illegible] Leu Arg Val
130 140

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Leu Asp Pro Ser Leu Val Ser Ser Leu Leu Asn Ala Val Ser Cys Gly
20 25 30

Ala Glu Ser Ser Ala Ser Ser Ser Ala Arg Ser Gly Asn Gly Ser Arg
35 40 45

Trp Thr Ser Met Pro Ser Gly Thr Arg Pro Gly Pro Arg Arg Ala Thr
50 55 60

Ser Arg Asp Asp Arg Arg Ser Ala Thr Ser Val Ile Pro Ser Arg Arg
65 70 75 80

Ser Val Ala Pro Arg Ala Glu Pro Gly Thr Arg Leu Ala Ser His Arg
85 90 95

Ala Ser Leu Ser Asn Ala Cys Ile Val Arg Leu Val Ile Ser Ala Ser
100 105 110

Gly Arg Ile Leu Ser Ser Leu Ile Leu Thr Arg Ser Arg Cys Val
115 120 125

Asp Lys Asn Gly Arg Arg Cys Ala Ser Gly Cys Arg Leu Asn Arg
130 135 140

Ala Arg Ser Ser Ser Ile Ala Asp Ser Cys Arg Thr Ile Gly Thr Phe
145 150 155 160

Arg Arg Ser Arg Tyr Cys Ala Ser Thr Arg Val Ser Thr Asn Ser Ile
165 170 175

His Val Ser His Gly Val Ala Cys Gly Ser Ala Leu Ile Gly Gly
180 185 190

35 40 45

Arg Asn Pro Arg Arg Ser Ser Arg Arg Asp Ala Glu Asp Arg Arg Val
50 55 60

Ile Phe Ala Ala Thr Leu Val Ala Val Asp Pro Pro Leu Arg Gly Ala
65 70 75 80

Gly Gly Glu Ala Asn Gln Leu Ile Asp Leu Gly Val Cys Arg Arg Glu
85 90 95

Ala Gly Arg Val Arg Arg Gly His Glu Leu His His Arg His Arg His
100 105 110

Gln Gly Ala Ala [illegible] Asp [illegible] Arg Arg Arg Arg Arg His Arg Arg Val
115 120 125

Gln Gln His Arg Arg [illegible] Gln Arg Gln Leu Arg Arg Tyr Val
130 135 140

Glu Thr Ala His His Arg Arg Pro Ala Arg Thr Asp Arg Val Arg His
145 150 155 160

His Val Arg Gly Pro Ser Asn His Arg Arg Arg Arg Val Tyr Arg Gly
165 170 175

Arg His Ser Gly Arg Gly Thr [illegible] Ala Gly Gly Ala Gly Ser Val
180 185 190

Gly Gly Ser Ala
195

[illegible]

Ala Arg Gly Gly Lys Asp Ile Pro Ile Leu Leu Arg Val Thr Pro Gly
50 55 60

Ala Thr Pro Trp Val Thr Trp Gly Glu Phe Val Glu Thr Arg Met Leu
65 70 75 80

Ala Glu Tyr Arg Asp Arg Arg Lys Val Pro Ile Val Arg Gln Arg Ala
85 90 95

Ala Ile Glu Ala Leu Arg Ala Arg Ile Asn Leu Arg Tyr Pro Leu Ala
100 105 110

His Leu Arg Ile Ile Leu Leu Ser Thr His Glu Arg Asp Leu Thr Met Gly
115 120 125

Lys Glu Glu Leu Leu Leu Ile Asp Ala Tyr Val Thr Ile Arg Thr Gly
130 135 140

Ile Ala Leu Leu Lys Arg Arg Ala Leu Leu Val Pro Asn
145 150 155 160

Ser Ala Arg Gly Ala Thr Leu Arg Arg Leu Gly Leu Thr Asp Val Ala
165 170 175

Asp Leu Arg Pro Pro Arg Glu Thr Arg Arg Arg Gly Pro Gly Arg Val
180 185 190

Tyr Asp Gly Leu Asp Val Ile Leu Leu Pro Phe Phe Arg Leu Ala Asp
195 200 205

Asp Arg Arg Asp Asp Leu Ala Thr Ala Phe Lys Arg Leu
210 215 220

Phe Leu Leu Arg Gly Leu Val Asp Ala Ile Arg Ala Ala Arg Ala Ala
225 230 235 240

Val Asp Arg Ala Ala Ile Arg Met Ile Gly Ala Ala Ala
245 250 255

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

CTCGTGCCGA TTCGGCACGA GCTGAGCAGC CTAAGGGGCC GTTCGGCGAA GTCATCGAGG 60

CATT[illegible]GCGGA [illegible] GGCAAGGGTA AGTAAATCAA CACCACGG[illegible] AACAGCCTGT 120

[illegible] GAAC[illegible] AATGA[illegible] [illegible]GCCGACTT GTTCGGC[illegible] GTACGCAGCC 180

TGGCGCTATT [illegible] CTACATCAGG [illegible] GTTCCTCG[illegible] TTGAACAAGA 240

[illegible] [illegible] [illegible] [illegible] [illegible] AA[illegible]CCATCC [illegible]

AGTAATTCGA CAGCTTGCTT [illegible] [illegible] [illegible]GCCAAGAAC CG[illegible]GAGGTGC 360

[illegible] [illegible] [illegible] [illegible] TTGCTGCAG [illegible]GATCCGT 420

[illegible] GGAGACCGT[illegible] CTGCACATCT [illegible] [illegible] ATTAACCAGC 480

[illegible] [illegible] [illegible] [illegible] GTTCACGAAT [illegible]CCAACC 540

[illegible] [illegible] TGGATTCAG[illegible] [illegible] [illegible] CAAGACTCGC 600

[illegible] 660

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

TTGGCTCCTT CTTACTGCTA CAGAGTAAAT TTAAAAGGTT TTGCACTAA TAACGATGG 1440

AATCGGGGGT GCATATCATC CCACAGGGCC TTGGTGCCAT GCTGGCGATG CCGATCGCCG 1500

GAGCGATGAT GGACCGACGG GGACCGGCCA AGATCGTGCT GGTTGGGATC ATGCTGATCG 1560

CTGCGGGGTT GGGCACCTTC GCCTTTGGTG TCGCGCGGCA AGCGGACTAC TTACCCATTC 1620

TGCCGACCGG GCTGGCAATC ATTGCCATGG GCATAGGTTG CTCGATGATA TCACTGTCCG 1680

GCTCGGCAGT GACACCCTG CTGACAAT AGTCGCTTCC CGTTTGGATT CTGATCAGCG 1740

TCAACCAGCA GTTGGTTGT TCGATAGCCA CTTGATGAT TGTTTTTT CTCACCTACC 1800

ACTTTAATCA [illegible] ATTTTACTT CAAAAAACGT CGCATTAA TAAGAGAGTG 1860

[illegible] AACTTTGCGG 1920

CCTAACTGGT CTATGACGTT TCGCACGCCT ACGCTGTGCT ATTCGTTGCA CTGATTGCGC 1980

TACTGGTCTT GACGCTGAAT CCGTCGTTAT TCTCTCCGAA AAAGAAGCGT AGTCATCGAA 2040

CAGCACACTT CTATCTCCCA TGACTTTACC TT 2072

(2) INFORMATION FOR SEQ ID NO: [illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 197 base pairs
(B) TYPE: [illegible] acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

[illegible]

[illegible]	540
GAACATCGCC ACTGCAATCA CAACCAGCAG ATTTCCCACG CATACCCTTT CGTACCGGTG	600
CGCCGCGGTT GGTCGATCGG TCGCATATCG ATGGCGCCGT TTAACGTAAC AGCTTTCGCG	660
GGACCGGGGG TCACAACGGG CGAGTTGTCC GGCCGGGAAC CCGGCAGGTC TCGGCCGCGG	720
TCACCCCAGC TCACTGGTGC ACCATCCGGG TGTCGGTGAG CGTGCAACTC AAACACACTC	780
AACGGCAACG GTTTTTCAGG TTATCAGTTT AACCTCGATC CGCAATCGCT CGTACGTTTC	840
[illegible]	900
[illegible]	960
[illegible]	1020
[illegible]	1080
[illegible]	1140
[illegible]	1200
[illegible]	1260
[illegible]	1320
[illegible]	1380
[illegible]	1440
[illegible]	1500
[illegible]	1560
[illegible]	1620
[illegible]	1680

(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

CTGGCGTGCC AGTGTCACCG GCGATATGAC [illegible] AATTTCGCGG [illegible]CCGCCGGA 60

[illegible]GCCA [illegible] ACCACC[illegible] [illegible]CAATT[illegible] [illegible]GAAGGTC[illegible] [illegible]CAAGTGCCT 120

[illegible]AATGTG GTG[illegible]TGGTT [illegible] [illegible] TAT[illegible] [illegible]TACCCCA 180

AACAA[illegible] [illegible]CAGGAAT [illegible] [illegible]GATT [illegible] TCTGCTAGCC 240

[illegible]GGATGGTT [illegible]AGACGTAA[illegible] GGTTGGCTAG [illegible] [illegible] [illegible] [illegible]

[illegible]ATGGCA [illegible]AAATTA[illegible] AAAAC[illegible] [illegible] [illegible]GATTG[illegible] [illegible] [illegible]

AGT[illegible]TTGGT [illegible]TTCGTCTT TA[illegible]CATT[illegible] [illegible] [illegible] [illegible]CTGGAA [illegible]

GGGGACAACG TGCTTTT[illegible] TCTTG[illegible] [illegible] [illegible] [illegible]TGGCGAAA 480

[illegible] [illegible]GAATC[illegible] [illegible] [illegible] [illegible]TGC[illegible] [illegible] [illegible]

[illegible]ACCGATG ATTAT[illegible]A [illegible] [illegible] [illegible] TCTCGATGTC 600

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

CCGGCTCGTT GTTGGCATAC TCCGCCGCGG CGGCCTCGAC CGCACTGGCC CTGGCGTGTG 60

TCCGGGCTGA CCACCGGGAT CGCCGAACCA TCCGAGATCA CCTCGCAATG ATCCACCTCG 120

CGCAGCTGGT CACCCAGCCA CGGGGCGGTG TGCGACAGCG CCTGCATCAC CTTGGTATAG 180

CCCTCGCGCC CCACCCGCAG GAACTTCTAG TACTCGCCCA CCACCTGGTT ACCCGGACGG 240

GAGAAGTTCA GGCTCAAGCT CGCCATGTTC CCTCCGAGCT AGTTGACCCG GAAAACCAGA 300

TCTTTTGGCA GGTGCTCGGG CCCCTGCCAC ACGACAAACC CGACGCCGGG ATAGGTCAG 359

(2) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

Gln Gln Pro Lys Gly Pro Phe Gly Gln Val Ile Gln Ala Phe Ala Asp
1 5 10 15

Gly Leu Ala Gly Lys Gly Lys Gln Ile Asn Thr Thr Leu Asn Ser Leu
20 25 30

Ser Gln Ala Leu Asn Ala Leu Asn Gln Gly Arg Gly Asp Phe Phe Ala
35 40 45

Val Val Asn Ser Leu Ala Leu Phe Val Asn Ala Leu His Gln Asp Asp
50 55 60

Ile Gln Phe Val Ala Leu Asn Lys Asn Leu Ala Gln Phe Thr Asp Arg
65 70 75 80

Leu Thr His Ile Asp Asp Asn Leu Ser Asn Ala Ile Leu Gln Phe Asp
85 90 95

Ser Leu Leu Ala Val Ala Arg Ile Phe Ala Lys Asn Arg Gln Val
100 105 110

Leu Thr His Asp Val Asn Asn Leu Ala Val Thr Thr Thr Leu Leu
115 120 125

Thr Pro Asp Ile Leu Asp Gly Leu Val Leu His Ile Phe Pro
130 135 140

Thr Leu Ala Ala Asn Asn Asn Leu Tyr His Leu Thr His Gly Gly
145 150 155 160

Val Thr Leu Ala Ile Thr Ala Asn Ile Ser Ala Phe
165 170

Val Ala Arg Ile Gly Ala Ile Val Pro Gln Leu Leu
180 185

Ser Leu Ala Gln Leu Met Gly Gly Pro Asp Ile Ala Pro Pro Ser Ser
290 295 300

Gly Leu Gln Thr Pro Pro Gly Pro Pro Asn Ala Tyr Asp Glu Tyr Pro
305 310 315 320

Val Leu Pro Pro Ile Gly Leu Gln Ala Pro Gln Val Pro Ile Pro Pro
325 330 335

Pro Pro Pro Gly Pro Asp Val Ile Pro Gly Pro Val Pro Pro Val Leu
340 345 350

Ala Ala Ile Val Phe Pro Arg Asn Arg Pro Ala Ala Ser Glu Asn Phe
355 360 365

Arg Tyr Met Gly Leu Leu Leu Leu Ser Pro Gly Leu Ala Thr Phe Leu
370 375 380

Phe Gly Val [illegible] [illegible] Ser Pro Ala Ala [illegible] Thr Met Ala Arg Arg His
385 390 395 400

Val Leu Ile Pro Ala Leu Thr Gly Leu Ala Leu Ile Ala Ala Phe Val
405 410 415

Ala His Ser Trp Tyr Arg Thr Gln His Pro Leu Ile Asp Met Arg Leu
420 425 430

Phe Gln Asn Arg Ala Val Ala Gln Ala Asn Met Thr Met Thr Val Leu
435 440 445

[illegible] Leu [illegible] Leu Phe [illegible] Ser [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]
450 455 460

595 600 605

Gln Phe Asn His Ser Glu Ile Ile Ala Thr Ala Lys Lys Val Ala Leu

610 615 620

Thr Pro Glu Ser Gly Ala Gly Arg Gly Ala Ala Val Asp Pro Ser Ser

625 630 635 640

Leu Pro Arg Gln Thr Asn Phe Ala Ala Gln Leu Leu His Asp Leu Ser

645 650 655

His Ala Lys Ala Val Phe Val Ile Ala Thr Ala Leu Val Val Ser

660 665 670

Thr Leu Ile Pro Ala Ala Ile Leu Ile Lys Gln Glu Ala Ser His Arg

675 680

Ala Ala Ile Leu Leu Ala

685

2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

[illegible]

Thr Arg Arg Asp Ile Arg Thr Ala
115 120

(2) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 89 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

Leu Ala Tyr Gln Tyr Ile Asn Asn Tyr Asp Val Gly Leu Gln Phe Arg
1 5 10 15

Gly Pro Ala Gly Ile Val Ala Thr Ile Ser Gly Pro Ile Gly Pro Ser
20 25 30

Ile Ala Gln Gly Arg Gln Val Asn Ala Thr Gly Gly Ala Gly Phe Leu
35 40 45

Gln Arg Asn Ile Ala Val Ser Gly Asn Ser Pro Ile Asn Asn Ser
50 55 60

Pro Gly Ile Asn Ser Arg Ala Asp Ser Gln Arg His Leu Leu Asp
65 70 75 80

Tyr Ser Gly Ser Asp Val Gln Val Gly

(2) INFORMATION FOR SEQ ID NO:192:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

His Leu Asn Met Ile His Leu Ala Gln Leu Val Thr Gln Pro Pro Gly
35 40 45

Gly Val Arg Gln Arg Leu His His Leu Gly Ile Ala Val Ala Pro Gln
50 55 60

Pro Gln Glu Val Val Val Leu Ala His His Leu Val Thr Gly Thr Gly
65 70 75 80

Glu Val Gln Gly Glu Gly Arg His Val Ala Ala Glu Val Val Asp Pro
85 90 95

Thr Asn Gln Ile Leu Ala Gln Val Leu Gly Pro Ala Pro His Asp Lys
100 105 110

Ile Arg Ala Ala Gly Leu Gly Gln
115

(2) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

Ala Ala Asn Gly Pro Ala Ala Ala Val Leu Gln Thr Ala Ala Asp Val
1 5 10 15

Val Thr Gly Gly Val Ala Ala Ala Ala Ala Ala Gly Ala Val Ala

(2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 811 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GTCTCGCGAT GTGGCCGAGT ATGACTTTGA GAAGACCAAG CGTAGTAATC GAAGATATCG 60

GACTTTGTGG TCCCGGTGGC GGGATAGAGC ACCTGTCAGC GTTGGTCAGC GTCACCCGTT 120

GCTCGGACGC CGAACCCATG CTTTCAACGT AGCCTGTAGG TCACACAAGT CGCGAGCGTA 180

ACGTCACGGT CAAATATCGC GTGGAATTTC GCCGTGACGT TCCGCTCGCG GACAATCAAG 240

GCATACTCAC TTACATCGGA GCAATTTGGA CGTTTGGATC CGTTTCCGCT CTGGTGAACG 300

TCCCGTCAAA GGTGTAATCT CTAACAGTAG ATAAGAGAAT TAAGTTCTAT CAGGTGCACG 360

CTAAAGACAA TGGACGCATC TCTTACAAGC CGCTCTGCGA GGCGTGTGGT GAGGTGGTCG 420

ATACAGCGA TTTCGCTTAC GTTTACGATC TTCTTGACTT TCAAATGTTT TCGATCACCG 480

ATTACGACAT AGTAGTTTCT TTAGAAAAA TTAATTCTGA ATCTGAGGTT TTTGAGTTCG 540

TCGTAGCCGT CGACTTCGAG CCGATGATCT TCGACTCTAG CTACTTTTTC GAGCCTGATT 600

CTGAGTCTTC TAAATCTATT TTCGTCGTCG CTAAAAATGT CTTGAGACCT AGCGGATGG 660

CTTTCGTCGA TTCGCTATCT TTCGTGAAAT CTATAAAAAA TAAGAGCCGT TATCCACAAT 720

TCGTCGTCGA CTTCGCTAGT GACAAACTTT ATGAATGATT AGATATTTCT GAGTTGAAGT 780

CTTTTTATAT ATGAATCTTT TTCTTGATT ATTAAAGTGA AAATAGTCTT ATTCATGCCA 840

TTGTCTTTTT TTACTAGTAG TAGTTTTTT TGTATTTTAT AGTTTTTTAT TTTTGATCTT 900

AATTTTTAAT TTATTTAATT TTTTTTGAGA ATTTTTTTTT TTTTTTTTTT TTTATTTTTT 960

ATTTT 965

CTCCGCCGCT GCCACCGTTG TCGCCGTTGC CGATCAACAT GCCGCTGGCG CCACCCTTGC 240

CACCCACGCC ACCGGCTCGG CCCACCCCGC CGACACCAAG CGAGCTGCCG CCGGAGCCAC 300

CATCACCACC TACGCCACCG ACCGCCCAGA CACCAGCGAC CGGGTCTTCG TGAAACGTCG 360

CGGTGCCACC ACCGCCGCCG TTACCGCCAA CCCCACCGGC AACGCCGGCG CCGCCATCCC 420

CGCCGGCCCC GGCGTTGCCG CCGTTGCCGC CGTTGCCGAA CAACAACCCG CCGGCGCCGC 480

CGTTGCCGCC CGCGCCGCCG [illegible] [illegible] GCCAAGCCCG CTGCCGCCCT 540

[illegible] [illegible] [illegible] CATCGGTTC TGCCTCGG[illegible] TCTGGGCTGT 600

[illegible] GATG[illegible] TTGCCGCC[illegible] [illegible]CCCGGG CCCCCCCGTG GCGCCGTCAC 660

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] 720

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]GCACTGG 780

CAGCCCCGTT A[illegible]TAC[illegible]GAAA [illegible]CGTTGCTAC CA[illegible]TAGA [illegible]GTCGCAGTG CCGATGTGTA 840

CGAAAGCGCC GCCTCCGGCG CCGCCGCTAC [illegible]CACT GCCGCCGGCT ACACCGTCGC 900

[illegible] [illegible] [illegible] [illegible] [illegible] A[illegible]CGT 960

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]GGCACCGC 1020

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]CACCGGTCG 1080

[illegible] 1140

[illegible]

[illegible]

[illegible]

CGGTCGTATT AGCCGTGTTC ACTTGAAAC TAATATTTA TAAACTTAAA TGTCAGTTTC 1860

CTAGCAGCTG CGCGCTCGCT ATCATTAAGT AATCTTTCTA CTTCTGGATA TCTTATATCG 1920

CCGCACCGTG TCCCCAGCGG CCACGTGACC TTTGGTCGCT GGCTGGCGGC CCTGACTATG 1980

GCCGCGACGG CCCTCGTTCT GATTCGCCCC GGCGCGCAGC TTGTTGCGCG AGTTGAAGAC 2040

GGGAGGACAG GCCGAGCTTG GTGTAGACGT GGGTCAAGTG CGAATGCATA CTCCGCGGCG 2100

ACATGAATAG GCGGACGCCG ATGTCCTTGT TCCTGAGTCC CTCACTGACC ACTAGAGCCA 2160

CCTCAAGCTC TCTCCCTCCT AACGCGCCCC AGCCACTTCT CCGCCTTTA TTCGACCGCC 2220

GGGCTCCTCG CGCTACTCGC ATCGCCTCAT CTATCGATAA CGCAGTTCCT TCGGCCCAGG 2280

CATCCTCCAA CTCTTTTCCA CCTATGACTC TTCCAAGGCT TCTTACTCTC ATTTACAGC 2341

TCTCTTCAA GACTCCAAAA CCCACC 2367

(2) INFORMATION FOR SEQ ID NO:1961:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 376 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1961:

Gly Gly Thr Val Ala Ala Gly Ala Thr Gly Arg Arg Ala Gly Ser Ala
115 120 125

Met Ala Ala Arg Ala Ala Val Ala Ala Gly Leu Ile Thr Asp Ala Gly
130 135 140

His Ile Cys Arg Ala Val Pro Gly Ala Gly Arg Gly Ala Gly Arg Gly
145 150 155 160

Ile Asp Pro Val Cys Pro Gly Glu Ala Gly Ala Ala Gly Thr Thr Gly
165 170 175

Ala Ala Met Ala Gln Gln Pro Gly Val Ala Ala Val Thr Ala Arg Thr
180 185 190

Pro Asp Ala Cys Gly His Ala Gly Ala Ala Asp Thr Ala Val Ala Ala
195 200 205

Val Ala Pro [illegible] Pro Pro Pro Val Pro Thr Gly Thr Ala Gly Arg Ala
210 215 220

Gly Thr Thr Gly Pro Ala Val Ala Ala Val Ala Asp Gln Pro Gly Arg
225 230 235 240

Ala Ser Ala Ala Ala Gly Leu Thr Gly Pro Ala Ser Arg Ala Val Ala
245 250 255

Thr Val Ala Cys Gln Gln Pro Ala Gly Arg Ala Arg Leu Pro Gly Cys
260 265 270

Arg Pro [illegible] [illegible] Ala Val [illegible] Asn [illegible] [illegible] Ala Pro [illegible] Cys Arg Leu

[illegible]

[illegible]

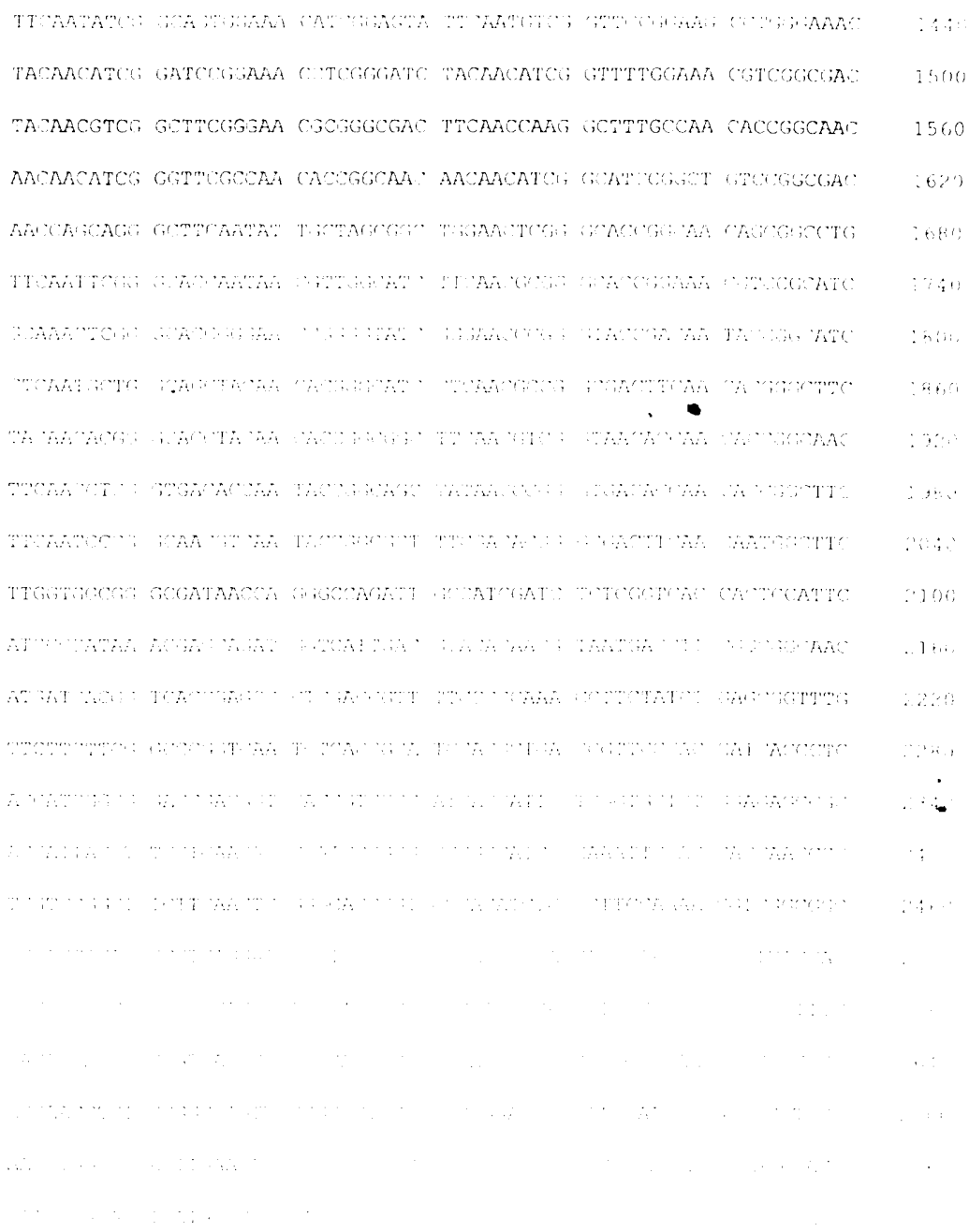
[illegible]

[illegible]

[illegible]

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198.



(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Gly Gln Asn Ala Pro Ala Ile Ala Ala Thr Glu Ala Ala Tyr Asp Gln
1 5 10 15

Met Trp Ala Glu Asp Val Ala Asn Met Ile Gly Tyr His Ala Gly Ala
20 25 30

Ser Ala Ala Val Ser Asn Leu Thr Ser Ile Gly Gln Ala Leu Leu Thr
35 40 45

Val Asn Gly Gly Gly Ala Leu Val Ala Ala Ala Gln Val Thr
50 55 60

Thr Asn Val Phe Asp Asn Leu Gly Leu Asn Val Asp Gln Gly Asn
65 70 75 80

Val Asn Asn Gly Asn Val Asn Asn Leu Gly Asn Ile
85 90 95

Gly Asn Gly Asn Leu Gly Ser Gly Asn Asn Ile Gly
100 105 110

Phe Gly Asn Val Gly Ile Gly Leu Ile Asn Asn Leu Asn Leu Gly
115 120 125

Ile Ala Thr Gly Ser Asn Asn Ile Gly Ser Gly Asn Ile
130 135 140

Gly Ser Tyr Asn Pro Gly Asn Ser Asn Thr Gly Gly Phe Asn Met Gly
245 250 255

Gln Tyr Asn Thr Gly Tyr Leu Asn Ser Gly Asn Tyr Asn Thr Gly Leu
260 265 270

Ala Asn Ser Gly Asn Val Asn Thr Gly Ala Phe Ile Thr Gly Asn Phe
275 280 285

Asn Asn Gly Phe Leu Trp Arg Gly Asp His Gln Gly Leu Ile Phe Gly
290 295 300

Ser Pro Gly Phe Phe Asn Ser Thr Ser Ala Pro Ser Ser Gly Phe Phe
305 310 315 320

Asn Ser Gly Ala Gly Ser Asn Ser Gly Phe Leu Asn Ser Gly Ala Asn
325 330 335

Asn Ser Gly Phe Phe Asn Ser Ser Ser Gly Ala Ile Gly Asn Ser Gly
340 345 350

Leu Ala Asn Ala Gly Val Leu Val Ser Gly Val Ile Asn Ser Gly Asn
355 360 365

Thr Val Ser Gly Leu Phe Asn Ser Ser Leu Val Ala Ile Thr Thr Pro
370 375 380

Ala Ser Ile Ser Gly Phe Phe Asn Thr Gly Ser Asn Met Ser Gly Phe
385 390 395 400

Phe Gly Ile Ser Ser Phe Asn Leu Gly Leu Ala Asn Arg Gly Val
405 410 415

[illegible]

[illegible]

530 [illegible] 540

Phe Asn Ile Ala Ser Gly Trp Asn Ser Gly Thr Gly Asn Ser Gly Leu
545 550 555 560

Phe Asn Ser Gly Thr Asn Asn Val Gly Ile Phe Asn Ala Gly Thr Gly
565 570 575

Asn Val Gly Ile Ala Asn Ser Gly Thr Gly Asn Trp Gly Ile Gly Asn
580 585 590

Pro Gly Thr Asn Asn Thr Gly Ile Leu Asn Ala Gly Ser Tyr Asn Thr
595 600 605

Gly Ile Leu Asn Ala Gly Asp Phe Asn Thr Gly Phe Tyr Asn Thr Gly
610 615 620

Ser Tyr Asn Thr Phe Gly Ile Asn Val Tyr Asn Thr Ala Thr Gly Asn
625 630 635 640

Ile Asn Val Gly Asp Thr Asn Thr Tyr Ser Ile Asn Phe Gly Asp Thr
645 650 655

Asn Thr Gly Phe Phe Asn Ile Tyr Asn Val Asn Thr Gly Ala Phe Asp
660 665 670

Thr Gly Asp Phe Asn Asn Gly Phe Leu Val Ala Thr Asn Gln Asn Gly
675 680 685

Gln Ile Asn Ile Asp Ile Ser Val Thr Thr Ile Ile Ile Asn Gly
690 695 700

[illegible]

[illegible]

[illegible]

Ala Ile Gly Asn Ser Gly Phe Gln Asn Leu Gly Ser Leu Gln Ser Gly
835 840 845

Trp Ala Asn Leu Gly Asn Ser Val Ser Gly Phe Phe Asn Thr Ser Thr
850 855 860

Val Asn Leu Ser Thr Pro Ala Asn Val Ser Gly Leu Asn Asn Ile Gly
865 870 875 880

Thr Asn Leu Ser Gly Val Phe Arg Gly Pro Thr Gly Thr Ile Phe Asn
885 890 895

Ala Gly Leu Arg Asn Leu Gly Gln Leu Asn Ile Gly Ser Ala Ser Cys
900 905 910

Arg Ile Asn His Thr Leu Asp Thr Val Ser Thr Ile Leu Ser Ala Ile
915 920 925

Cys Gly Ser Ala Ser Asp Gln Ser Asn Pro Gly Ser Val Ser Gln
930 935 940

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible]

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

GGATCCTCTA GAGTCGACCT GCAGGCATGC A 31

(2) INFORMATION FOR SEQ ID NO:263:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

TGCATGCCTG CAGGTCGACT CTAGAGGATC C 31

(2) INFORMATION FOR SEQ ID NO:264:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

GGATCCTCTA GAGTCGACCT GCAGGCATGC A 31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

CGATATCTGC AGAATTCAGC TTTAAAGCCC ATTTGCGA 38

(2) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

TCGATATCGA GCCACTTGCT TAAAACGCCT 30

(2) INFORMATION FOR SEQ ID NO:207:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC 120

CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TCAAGCTCTA AATCGGGGGC TCCCTTTAGG 180

GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCCAAAAAA CTTGATTAGG GTGATCGTTC 240

ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT 300

CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACACTC AACCCTATCT CGGTCTATTC 360

TTTTGATTTA TAAGGGATTT TGCCGATTTC GGCCTATTGG TTAAAAAATG AGCTGATTTA 420

ACAAAAATTT AACGCGAATT TTAACAAAAT ATTAACGTTT ACAATTTCAG GTGGCACTTT 480

TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA 540

TCCGTTCATG AATTAATTCT TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT 600

TCATATCAGG ATTATCAATA CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA 660

ACTCACCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC 720

GTCCAACATC AATACAACCT ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGTGAGA 780

AATCACCATG AGTGACGACT GAATCCGGTG AGAATGGCAA AAGTTTATGC ATTTCTTTCC 840

AGACTTGTTC AACAGGCCAG CCATTACGCT CGTCATCAAA ATCACTCGCA TCAACCAAAC 900

CGTTATTCAT TCGTGATTGC GCCTGAGCGA GACGAAATAC GCGATCGCTG TTAAAAGGAC 960

[illegible]

AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC 1740

AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG 1800

CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC 1860

ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA 1920

AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT 1980

CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG 2040

CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG 2100

GCCTTTTTAC GGTTCCTGGC CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA 2160

TCCCCTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC 2220

AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG 2280

TATTTTCTCC TTACGCATCT GTGCGGTATT TCACACCGCA TATATGGTGC ACTCTCAGTA 2340

CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGTATACACT CCGCTATCGC TACGTGACTG 2400

GGTCATGGCT GCGCCCCGAC ACCCGCCAAC ACCCGCTGAC GCGCCCTGAC GGGCTTGTCT 2460

GCTCCCGGCA TCCGCTTACA GACAAGCTGT GACCGTCTCC GGGAGCTGCA TGTGTCAGAG 2520

[illegible]

[illegible]	3420
CCGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGCATC GGTCGAGATC CCGGTGCCTA	3480
ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA	3540
CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT	3600
TGGGCGCCAG GGTGGTTTTT CTTTTCACCA GTGAGACGGG CAACAGCTGA TTGCCCTTCA	3660
CCGCCTGGCC CTGAGAGAGT TGCAGCAAGC GTCCACGCT GGTTTGCCCC AGCAGGCGAA	3720
AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTTC TTATCATGCT	3780
ATCCATTAT CGAGATTCTT GTAAAACCT CTAGCTCCGA CTCCGTAATT CCGCAATTG	3840
CGCCTACCGG CATCTGATCG TTGGCAACGA CCATCGAGT TTAACTATC TCTTATTGA	3900
CGATTTGCAT GCTTTGTTA AAACTTAGA TGCATTCGA GTTTTTTC TTTTCGCTA	3960
TCGCTTAAT TTGATTGCGA GTGAGATATT TATGCCAGT ATCTAGACTG AGACTCGCCG	4020
AGACAGAACT TAATGGGCCC GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT	4080
GCTCCACGCC CAGTCGCGTA CCGTCTTCAT GGGAGAAAAT AATACTGTTG ATCCGTGTCT	4140
CGTCAGAGAC ATCAAGAAAT AACGTCCGAA CATTAATTTA CGTACGTTCG ATCAATGG	4200
CATCCTCGCT ATTGATCAAA TAGTTAATGA CGATCTCAT GATCTTTTCG TCAAAAGAT	4260
[illegible]	432
[illegible]	4
[illegible]	4
[illegible]	4
[illegible]	4
[illegible]	4

AATTAATACG ACTCACTATA GGGGAATTGT GAGCGGATAA CAATTCCCCT CTAGAAATAA 5040

TTTTGTTTAA CTTTAAGAAG GAGATATACA TATGGGCCAT CATCATCATC ATCACGTGAT 5100

CGACATCATC GGGACCAGCC CCACATCCTG GGAACAGGCG GCGGCGGAGG CGGTCCAGCG 5160

GGCGCGGCAT AGCGTCGATG ACATCCGCGT CGCTCGGGTC ATTGAGCAGG ACATGGCCGT 5220

GGACAGCGCC GGCAACATCA CCTACCGCAT CAAGCTCGAA CTGTCGTTTA AGATGAGGCC 5280

GGCGCAACGG AGGGGCTGGA AACCACCGAT CGGTTTGCCT GAAACGGGGG CCGGCGCCGG 5340

TAGTGTGGGG ACTACCCTCG CGTCTTGGCC TTTTGCTTTG GCGGAGACCG GTAGCACCCT 5400

GCTCTACCTG CTGTTCAACC TGTGGGGTCC GCGTTTTCAC GAGAGGTATC CGAACGTCAC 5460

GATCACCGCT CACGGGCATC TTTTTGTCGC CGGATCGCAC CGCGCCGAGC CCGGGACGGT 5520

CAACATTGGG CGTCTGAACG CCTATTCTCG GGAAGTCGAT ATTAAGGTTA ACAAGGGCCT 5580

GATGAACAGC GCCGTAGCGC TCCCGTTCA GAATTGCAAC TACAACCTCA CCGGAGTGAG 5640

CGAGCACCTC AAGCTGAACG GAAAAGTCCT GAAGTCGATC TACCACGGCA CCATCAAAAC 5700

CTCGGATGAC CGGAGATTCG TTGCGATCAA GTCCGACCTG AATCTGGCTA CGACGCGGT 5760

AGTTCCGCTG CATCTCTCCA ACGACTCGAC TTACACTTTC TTTTTCACTC AGTACGTGTC 5820

CAAGGAAGAT GCCATAGCCT CGGTAAGTC CTCAGTTTT GCAGCAACG TCGACTTCCC 5880

CCTCGTCGCC GTTCTCCTGC CGAGAACTCC CAAGGACTTC ATTCCGATCA CTCTGACCAG 5940

GACATCCGTC TCCGAGTCAT ACATCGAGTG GCAGCAGACC ATCCAGAGGA TCGCAGAGTC 6000

ATGGCTCTCA CCGCTCGAGC ACCACCACCA CCACCACTGA GATCCGGCTG CTAACAAAGC 6060

CCGAAAGGAA GCTGAGTTGG CTGCTGCCAC CGCTGAGCAA TAACTAGCAT AACCCCTTGG 6120

GGCCTCTAAA CGGGTCTTGA GGGGTTTTTT GCATGAACCA TAAGATGACT TTGTCATGAG 6180

CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCCGC GCACATTTCC 6240

CCGAAAAGTG CCACCTGACG TCTAAGAAAC CATTATTATC ATGACATTAA CCTATAAAAA 6300

TAGGCGTATC ACGAGGCCCT TTCGTCTTCA AGAATTCTCA TGTTTGACAG CTTATCATCG 6360

ATAAGCTTTA ATGCGGTAGT TTATCACAGT TAAATTGCTA ACGCAGTCAG GCACCGTGTA 6420

TGAAATCTAA CAATGCGCTC ATCGTCATCC TCGGCACCGT CACCCTGGAT GCTGTAGGCA 6480

TAGGCTTGGT TATGCCGGTA CTGCCGGGCC TCTTGCGGGA TATCGTCCAT TCCGACAGCA 6540

TCGCCAGTCA CTATGGCGTG CTGCTAGCGC TATATGCGTT GATGCAATTT CTATGCGCAC 6600

CCGTTCTCGG AGCACTGTCC GACCGCTTTG CCGATCGCCA CGTCGAGTCC TGACGCGTAA 6660

TCGCTACTTA TCTACGTGTC TGCGGTATTT CACACCGCAT ATGGTGCACT CTCAGTACAA 6720

TCTGCTCTGA TGCCGCATAG TTAAGCCAGC CCCGACACCC GCCAACACCC GCTGACGCGC 6780

CCTGACGGGC TTGTCTGCTC CCGGCATCCG CTTACAGACA AGCTGTGACC GTCTCCGGGA 6840

GCTGCATGTG TCAGAGGTTT TCACCGTCAT CACCGAAACG CGCGAGGCAG CAGATCAATT 6900

CGCGCGCGAA GGCGAAGCGG CATGCATTTA CGTTGACACC ATCGAATGGT GCAAAACCTT 6960

TCGCGGTATG GCATGATAGC GCCCGGAAGA GAGTCAATTC AGGGTGGTGA ATATGAAACC 7020

AGCAACTTTC TTCTCCTCAG GAGTCAGGAT TTCCACGATG CTCCTAGCGA TGAGCTTTAC 7080

CAAGAGATAT TTGATCTGCT GAATGCGCAA TCACGTGCTT ACGTGGATAC GGGCTTTAAT 7140

TCGCTGGAAA TCAGTCCACA GATTACGCAA GTGCCACTGG AGAAGGAGGA TAACCTGCAC 7200

AAGCGTCCGT TCGCCGTGCG CAAGCAGGAA ACGCGAGACG CCGTGGCCGG TGCCCGCGAT 7260

CTGGCCAGCG GCGTGAGCTT CATCGGTGAG CTGCCGATCG ACCCGCTGCC GCTGTACAAA 7320

AGCGACGGCA CGGTGACCTC GCTGCGCGCC AACGACGCCG AAGTGACCGG CAGCATGTCC 7380

TTCCTCGACG TGGAGCTGGC CACCCCGATC AAGAAGCAGT CCCTGCCGGT GTTCTGGCTG 7440

GACGCCAAGG GCGAGTACGT GCCGCTGTGG ATCAAGACCA GCCTGCTGAG CGTGCACAAG 7500

GGCGCCATCG CCTACCACGT GCCGCCGGTG CAGGCCGTGG TGAAGGACGG CGTGCTGGTG 7560

GAGCCGGCCA ACCTGACCAG CCGCAACAAG CCCAGCACCG AGGAGCTGGC CCACGAGCTG 7620

GAGTTCGACG CCCAGTGGCT GCCGGCCTTC GAGATCGACG CCGAGAGCGC CTTCAAGTAC 7680

GACTTCAAGG ACGGCGTGGC CATCCTGCCG ATCGGCCTGC AGGAGATCAA CGACCGCGAG 7740

ACCCGCATCC TGCTGCCGCT GAAGGAGCCC GACGGCCTGG CCGCCGCCGT GGCCATCCAC 7800

GGCGGCAAGG TGGAGGCCCG CGCCGTGGAG TTCCTGCCGG GCGAGGGCAA GCTGCTGCTG 7860

GGCGTGCCGC TGAGCGAGTA CGCCGGCCTG ACCCCGGCCA CCTTCCTGAA GCTGTTCGGC 7920

AGCCCGGTGG AGGTGGCCAA GGTGGAGGCC GGCGCCAGCG GCCCGAGCAG CGCCCAGGTG 7980

GTGAAGGCCT TCGAGAGCGC CAGCGAGAAG AAGCAGCTGC CGGTGTTCCT GCTGGACGCC 8040

AAGGGCGAGT ACGTGCCGCT GTGGATCAAG ACCAGCCTGC TGAGCGTGCA CAAGGGCGCC 8100

ATCGCCTACC ACGTGCCGCC GGTGCAGGCC GTGGTGAAGG ACGGCGTGCT GGTGGAGCCG 8160

GCCAACCTGA CCAGCCGCAA CAAGCCCAGC ACCGAGGAGC TGGCCCACGA GCTGGAGTTC 8220

GACGCCCAGT GGCTGCCGGC CTTCGAGATC GACGCCGAGA GCGCCTTCAA GTACGACTTC 8280

AAGGACGGCG TGGCCATCCT GCCGATCGGC CTGCAGGAGA TCAACGACCG CGAGACCCGC 8340

ATCCTGCTGC CGCTGAAGGA GCCCGACGGC CTGGCCGCCG CCGTGGCCAT CCACGGCGGC 8400

AAGGTGGAGG CCCGCGCCGT GGAGTTCCTG CCGGGCGAGG GCAAGCTGCT GCTGGGCGTG 8460

CCGCTGAGCG AGTACGCCGG CCTGACCCCG GCCACCTTCC TGAAGCTGTT CGGCAGCCCG 8520

GTGGAGGTGG CCAAGGTGGA GGCCGGCGCC AGCGGCCCGA GCAGCGCCCA GGTGGTGAAG 8580

GCCTTCGAGA GCGCCAGCGA GAAGAAGCAG CTGCCGGTGT TCCTGCTGGA CGCCAAGGGC 8640

GAGTACGTGC CGCTGTGGAT CAAGACCAGC CTGCTGAGCG TGCACAAGGG CGCCATCGCC 8700

TACCACGTGC CGCCGGTGCA GGCCGTGGTG AAGGACGGCG TGCTGGTGGA GCCGGCCAAC 8760

CTGACCAGCC GCAACAAGCC CAGCACCGAG GAGCTGGCCC ACGAGCTGGA GTTCGACGCC 8820

CAGTGGCTGC CGGCCTTCGA GATCGACGCC GAGAGCGCCT TCAAGTACGA CTTCAAGGAC 8880

GGCGTGGCCA TCCTGCCGAT CGGCCTGCAG GAGATCAACG ACCGCGAGAC CCGCATCCTG 8940

CTGCCGCTGA AGGAGCCCGA CGGCCTGGCC GCCGCCGTGG CCATCCACGG CGGCAAGGTG 9000

GAGGCCCGCG CCGTGGAGTT CCTGCCGGGC GAGGGCAAGC TGCTGCTGGG CGTGCCGCTG 9060

AGCGAGTACG CCGGCCTGAC CCCGGCCACC TTCCTGAAGC TGTTCGGCAG CCCGGTGGAG 9120

GTGGCCAAGG TGGAGGCCGG CGCCAGCGGC CCGAGCAGCG CCCAGGTGGT GAAGGCCTTC 9180

GAGAGCGCCA GCGAGAAGAA GCAGCTGCCG GTGTTCCTGC TGGACGCCAA GGGCGAGTAC 9240

GTGCCGCTGT GGATCAAGAC CAGCCTGCTG AGCGTGCACA AGGGCGCCAT CGCCTACCAC 9300

GTGCCGCCGG TGCAGGCCGT GGTGAAGGAC GGCGTGCTGG TGGAGCCGGC CAACCTGACC 9360

AGCCGCAACA AGCCCAGCAC CGAGGAGCTG GCCCACGAGC TGGAGTTCGA CGCCCAGTGG 9420

CTGCCGGCCT TCGAGATCGA CGCCGAGAGC GCCTTCAAGT ACGACTTCAA GGACGGCGTG 9480

GCCATCCTGC CGATCGGCCT GCAGGAGATC AACGACCGCG AGACCCGCAT CCTGCTGCCG 9540

CTGAAGGAGC CCGACGGCCT GGCCGCCGCC GTGGCCATCC ACGGCGGCAA GGTGGAGGCC 9600

CGCGCCGTGG AGTTCCTGCC GGGCGAGGGC AAGCTGCTGC TGGGCGTGCC GCTGAGCGAG 9660

TACGCCGGCC TGACCCCGGC CACCTTCCTG AAGCTGTTCG GCAGCCCGGT GGAGGTGGCC 9720

AAGGTGGAGG CCGGCGCCAG CGGCCCGAGC AGCGCCCAGG TGGTGAAGGC CTTCGAGAGC 9780

GCCAGCGAGA AGAAGCAGCT GCCGGTGTTC CTGCTGGACG CCAAGGGCGA GTACGTGCCG 9840

CTGTGGATCA AGACCAGCCT GCTGAGCGTG CACAAGGGCG CCATCGCCTA CCACGTGCCG 9900

CCGGTGCAGG CCGTGGTGAA GGACGGCGTG CTGGTGGAGC CGGCCAACCT GACCAGCCGC 9960

AACAAGCCCA GCACCGAGGA GCTGGCCCAC GAGCTGGAGT TCGACGCCCA GTGGCTGCCG 10020

GCCTTCGAGA TCGACGCCGA GAGCGCCTTC AAGTACGACT TCAAGGACGG CGTGGCCATC 10080

CTGCCGATCG GCCTGCAGGA GATCAACGAC CGCGAGACCC GCATCCTGCT GCCGCTGAAG 10140

GAGCCCGACG GCCTGGCCGC CGCCGTGGCC ATCCACGGCG GCAAGGTGGA GGCCCGCGCC 10200

GTGGAGTTCC TGCCGGGCGA GGGCAAGCTG CTGCTGGGCG TGCCGCTGAG CGAGTACGCC 10260

GGCCTGACCC CGGCCACCTT CCTGAAGCTG TTCGGCAGCC CGGTGGAGGT GGCCAAGGTG 10320

GAGGCCGGCG CCAGCGGCCC GAGCAGCGCC CAGGTGGTGA AGGCCTTCGA GAGCGCCAGC 10380

GAGAAGAAGC AGCTGCCGGT GTTCCTGCTG GACGCCAAGG GCGAGTACGT GCCGCTGTGG 10440

ATCAAGACCA GCCTGCTGAG CGTGCACAAG GGCGCCATCG CCTACCACGT GCCGCCGGTG 10500

CAGGCCGTGG TGAAGGACGG CGTGCTGGTG GAGCCGGCCA ACCTGACCAG CCGCAACAAG 10560

CCCAGCACCG AGGAGCTGGC CCACGAGCTG GAGTTCGACG CCCAGTGGCT GCCGGCCTTC 10620

GAGATCGACG CCGAGAGCGC CTTCAAGTAC GACTTCAAGG ACGGCGTGGC CATCCTGCCG 10680

ATCGGCCTGC AGGAGATCAA CGACCGCGAG ACCCGCATCC TGCTGCCGCT GAAGGAGCCC 10740

GACGGCCTGG CCGCCGCCGT GGCCATCCAC GGCGGCAAGG TGGAGGCCCG CGCCGTGGAG 10800

TTCCTGCCGG GCGAGGGCAA GCTGCTGCTG GGCGTGCCGC TGAGCGAGTA CGCCGGCCTG 10860

ACCCCGGCCA CCTTCCTGAA GCTGTTCGGC AGCCCGGTGG AGGTGGCCAA GGTGGAGGCC 10920

GGCGCCAGCG GCCCGAGCAG CGCCCAGGTG GTGAAGGCCT TCGAGAGCGC CAGCGAGAAG 10980

AAGCAGCTGC CGGTGTTCCT GCTGGACGCC AAGGGCGAGT ACGTGCCGCT GTGGATCAAG 11040

ACCAGCCTGC TGAGCGTGCA CAAGGGCGCC ATCGCCTACC ACGTGCCGCC GGTGCAGGCC 11100

GGCGGTGTCG ACTGGTTCAG CTGCAATTTT ATTGTCGACA CTTGTTCTTT CCCTACGACC 6720

GGCCGCCGCC AACACGTCGA ATGTCCAGCC GGGCGATCCC AACGCAGCAC CTCCGCCGGC 6780

CGACCCGAAC GCACCGCCGC CACCTGTCAT TGCCCCAAAC GCACCCCAAC CTGTCCGGAT 6840

CGACAACCCG GTTGGAGGAT TCAGCTTCGC GCTGCCTGCT GGCTGGGTGG AGTCTGACGC 6900

CGCCCACTTC GACTACGTGT CAGCACTCCT CAGCAAAACC ACCGGGGACC CGCCACTTCC 6960

CGAACAGCCG CCTCCTGTCG CAATGACAAC TTTATCGTTG CTGGCCCGGT TAGACTAAAA 7020

GCTTTAAGGC AGCGCCAAAG TCAGCGACTT CAAGATCGCG GTCCGGTTTG GTCGGACAT 7080

TGGTGATTTC TATATTTCCT AACCGGCAC TTGATGAA TA GAAAAACG TCTCGCTTGA 7140

CGTTAACGCG CTTTTTGGAA CTGCCTCTTA TTATTAATTT AACTTCTTTC ATTCGAGTAA 7200

TCGGAACGGC CATATCTTGA CGGGCGTAAT CTTTTTGTTT CTTCGGAACC TAATCGACGC 7260

TGCTGCCGCT CATCTCTGTT TTGTCGTATG CTTTGAATCT TTTAACTAAC CATTCGACAA 7320

GGCTGCCGCC AAAGCGCCTG CTGAATCGAT TTTGCTTTTT TTTTGCGCTT CTTCGTCGCC 7380

GCTAACGCTT TTTGTAGAAT TTTTTCTTTT TTTTGGGGAA TCGTTCTTAC 7440

CTTGACTACA CGGACTACGG AATCGACTTT AATGTTTTGA CAATTCTTTA CATATTCATC 7500

TATTTTGCTT TTTTTAACT ACTACGATTA TACTACTTA CATCTTCTTT TTAATAAAGC 7560

TTTAAATAAA TTTAATTTTT TTTATTAAT TTTTAATTAA CAATTAATAT AATTTTTTTG 7620

GTATTTTAAA TTTTTTTTA TTTTTTTTT TTTTTTTTTT TTTTTA 7666

Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
35 40 45

Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Glu Val
50 55 60

Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser
65 70 75 80

Gly Ser Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Asn Thr Thr Pro
85 90 95

Ala Ser Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr
100 105 110

Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Gly Arg Tyr Pro Asn
115 120 125

Val Thr Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln
130 135 140

Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser
145 150 155 160

Glu Gly Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala
165 170 175

Ile Ser Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser His His
180 185 190

Leu Leu Leu Asn Gly Lys Thr Leu Ala Asn Met Arg Gly Gly Thr Ile
195 200

Tyr Thr Trp Asp Asp Pro Lys Ile Ala Ala Asp Asn Pro Gly Val Asn
210 215

Leu Pro Asp Thr Ala Val Val Pro Leu Gln Arg Ser Asp Gly Ser Gly

Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp
340 345 350

Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile
355 360 365

Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala
370 375 380

Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp
385 390 395 400

Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp
405 410 415

Ala Leu Ile Ala Thr Ile Ser Ser Ala [illegible] Met Tyr Thr Asp Ala Ala
420 425 430

Thr Leu Ala Asn Glu Ala Gly Asn Phe Gln Arg Ile Ser Gly Asp Leu
435 440 445

Lys Thr Gln Ile Asp Gln Val [illegible] Ser Thr Ala Gly Ser Leu Gln Gly
450 455 460

Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
465 470 475 480

[illegible] Gln [illegible] Arg Ala Asn Tyr [illegible] Lys Gln [illegible] Leu Asp [illegible] [illegible] Ser
485 490 495

[illegible]

[illegible]

[illegible]

[illegible]

Gly Asp Pro Pro Phe Pro Gly Gln Pro Pro Pro Val Ala Asn Asp Thr
625 630 635 640

Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu
645 650 655

Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu
660 665 670

Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser
675 680 685

Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Pro Pro Tyr Tyr Glu Val Lys
690 695 700

Ile Ser Asp Pro Ser Lys Pro Asn Gly Gln Ile Trp Thr Gly Val Ile
705 710 715 720

Gly Pro Pro Ala Ala Asn Ala Ile Asp Ala Gly Pro Pro Gln Trp
725 730 735

Phe Val Val Trp Leu Gly Thr Asn Asn Asn Pro Val Asp Lys Gly Ala
740 745 750

Ala Lys Asn Val Ala Glu Pro Val Asn Arg Ile Leu Val Asn Pro
755 760 765

Ala Pro Ala Pro Ala Pro Ala Thr Pro Ala Pro Ala Pro Ala
770 775 780

Gly Leu Val Ala Leu Thr Leu Ala Leu Leu Ala
785 790 795

Ile Ala

CLAIMS

We claim:

1. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

(a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);

(b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala Leu Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);

(c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 17);

(d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);

(e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);

(f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);

(g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);

(i) Asp Pro Ala Ser-Ala Pro Asp Val Pro-Thr-Ala-Ala-Gln-Leu Thr-Ser Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn Val Ser Phe Ala-Asn (SEQ ID NO: 123); and

(j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131)

wherein Xaa may be any amino acid

2. A polypeptide comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

(a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124) and

(b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any amino acid.

3. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

4. A polypeptide comprising an antigenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196 or a complement thereof under moderately stringent conditions.

5. A DNA molecule comprising a nucleotide sequence encoding a polypeptide according to any one of claims 1-4

6. A recombinant expression vector comprising a DNA molecule according to claim 5.

7. A host cell transformed with an expression vector according to claim 6.

8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.

9. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with one or more polypeptides according to any of claims 1-4; and

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

10. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

11. A method for detecting *M tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

12. The method of any one of claims 9-11 wherein step (a) additionally comprises contacting the biological sample with a 38 kD *M. tuberculosis* antigen and step (b) additionally comprises detecting in the sample the presence of antibodies that bind to the 38 kD *M. tuberculosis* antigen.

13. The method of any one of claims 9-11 wherein the polypeptide(s) are bound to a solid support.

14. The method of claim 13 wherein the solid support comprises nitrocellulose, latex or a plastic material.

15. The method of any one of claims 9-11 wherein the biological sample is selected from the group consisting of whole blood, serum, plasma, saliva, cerebrospinal fluid and urine.

16. The method of claim 15 wherein the biological sample is whole blood or serum.

17. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a DNA molecule according to claim 5; and

(b) detecting in the sample a DNA sequence that amplifies in the presence of the oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

18. The method of claim 17, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA molecule according to claim 5.

19. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a DNA sequence selected from the group consisting of SEQ ID NOS. 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

20. The method of claim 19, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

21. The method of claims 17 or 19 wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

22. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with one or more oligonucleotide probes specific for a DNA molecule according to claim 5; and

(b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

23. The method of claim 22 wherein the probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

24. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with one or more oligonucleotide probes specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

25. The method of claim 24 wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

26. The method of claims 22 or 24 wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

27. A method for detecting *M tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide according to any one of claims 1-4; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

28. A method for detecting *M tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

29. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

30. The method of any one of claims 27-29 wherein the binding agent is a monoclonal antibody.

31. The method of any one of claims 27-29 wherein the binding agent is a polyclonal antibody.

32. A diagnostic kit comprising:

(a) one or more polypeptides according to any of claims 1-4; and

(b) a detection reagent.

33. A diagnostic kit comprising:

(a) one or more polypeptides having an N-terminal sequence selected from

34. A diagnostic kit comprising:

(a) one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) a detection reagent.

35. The kit of any one of claims 32-34 wherein the polypeptide(s) are immobilized on a solid support.

36. The kit of claim 35 wherein the solid support comprises nitrocellulose, latex or a plastic material.

37. The kit of any one of claims 32-34 wherein the detection reagent comprises a reporter group conjugated to a binding agent.

38. The kit of claim 37 wherein the binding agent is selected from the group consisting of anti-immunoglobulins, Protein G, Protein A and lectins.

39. The kit of claim 37 wherein the reporter group is selected from the group consisting of radioisotopes, fluorescent groups, luminescent groups, enzymes, biotin and dye particles.

40. A diagnostic kit comprising at least two oligonucleotide primers, at least one of the oligonucleotide primers being specific for a DNA molecule according to claim 5.

41. A diagnostic kit according to claim 40, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA molecule according to claim 5.

42. A diagnostic kit comprising a at least two oligonucleotide primers, at least one of the primers being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

43. A diagnostic kit according to claim 42, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

44. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe being specific for a DNA molecule according to claim 5.

45. A kit according to claim 44, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

46. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

47. A kit according to claim 46, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

48. A monoclonal antibody that binds to a polypeptide according to any of

49. A polyclonal antibody that binds to a polypeptide according to any of claims 1-4.

50. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.

51. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6 (SEQ ID NO: 99).

52. A fusion protein comprising a polypeptide having an N-terminal sequence selected from the group of sequences provided in SEQ ID NOS: 129 and 130.

53. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and the *M. tuberculosis* antigen 38 kD (SEQ ID NO: 150).

54. A diagnostic kit comprising:

(a) one or more fusion proteins according to any one of claims 50-53; and

(b) a detection reagent.

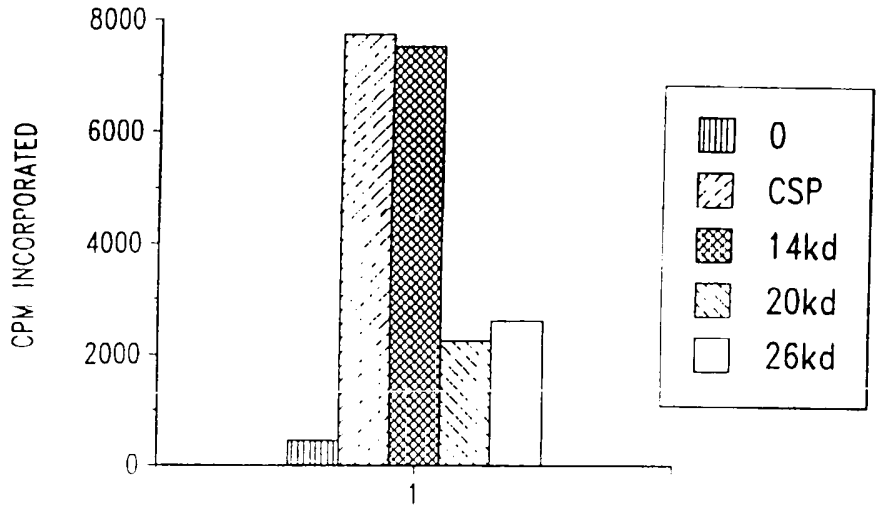


Fig. 1A-1

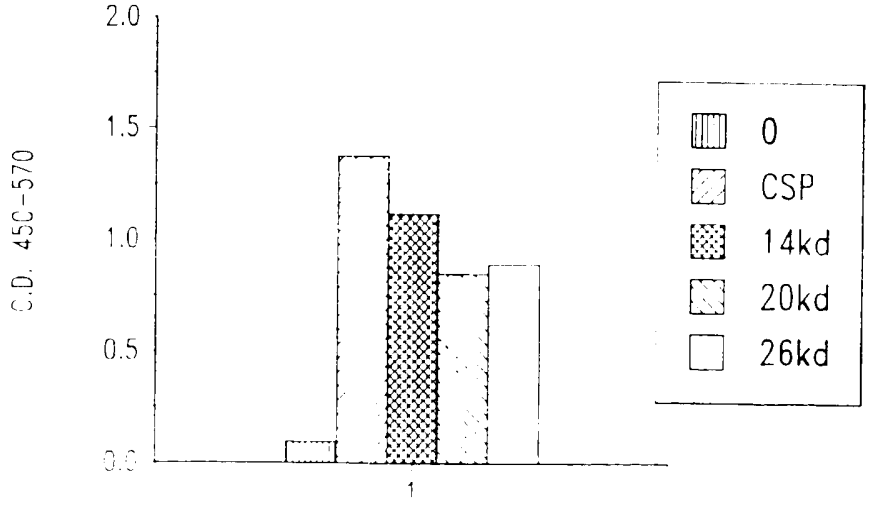


Fig. 1A-2

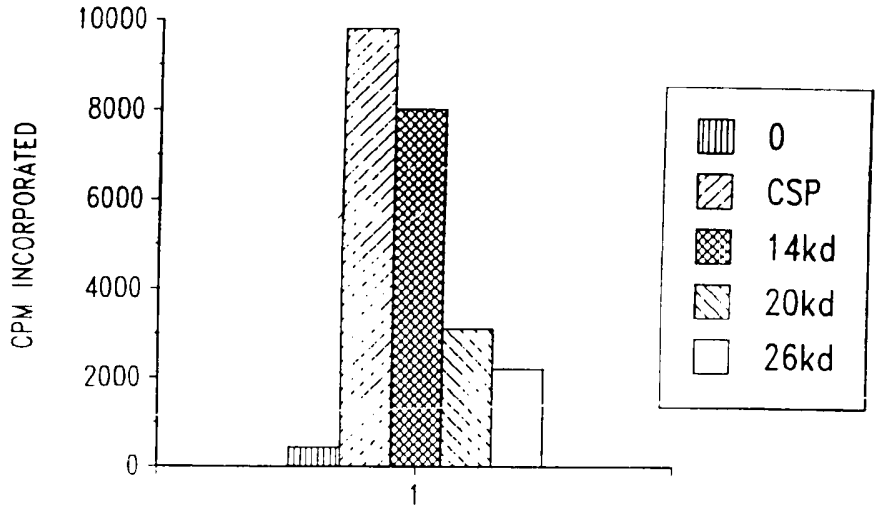

10000
8000
6000
4000
2000
0
CPM INCORPORATED
1
0
CSP
14kd
20kd
26kd

Fig. 1B-1

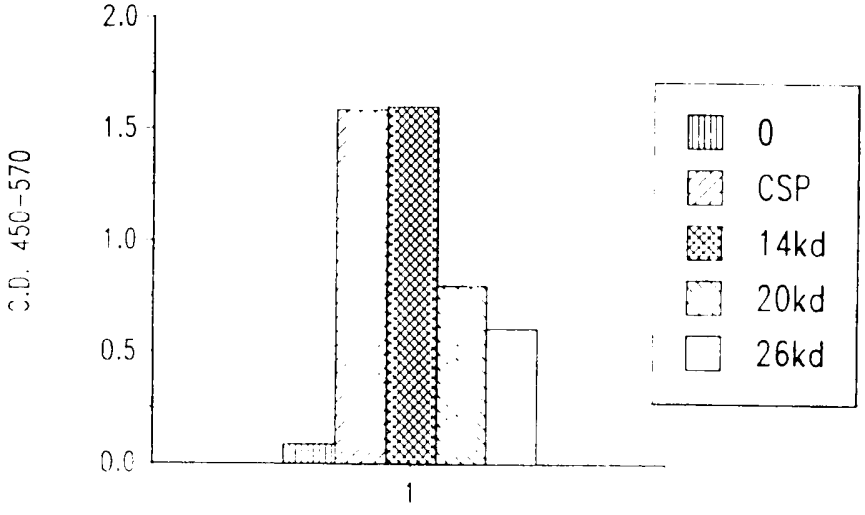

2.0
1.5
1.0
0.5
0.0
O.D. 450-570
1
0
CSP
14kd
20kd
26kd

Fig. 2B

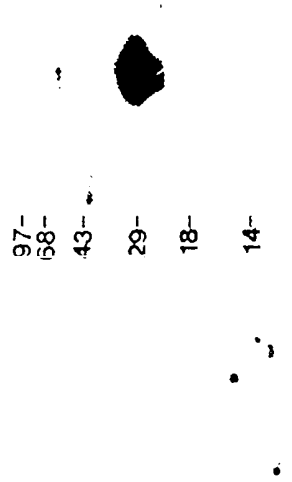
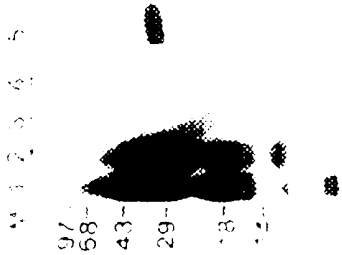
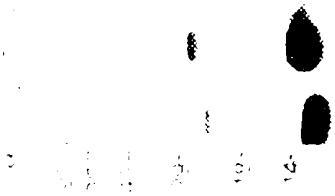


Fig. 2D





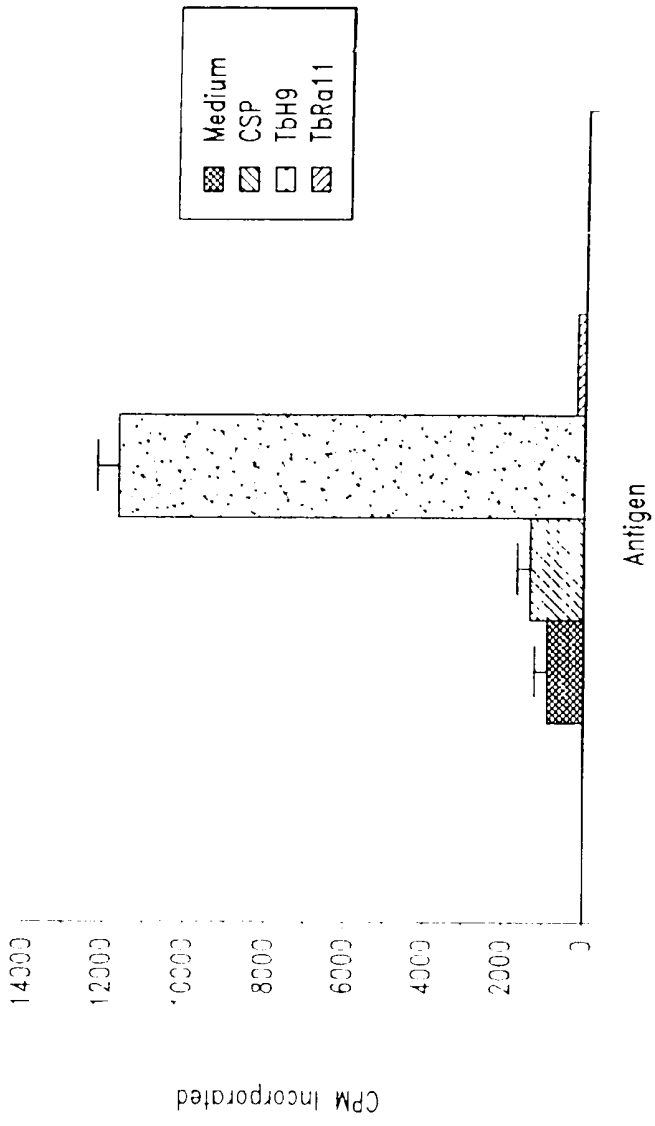


Fig. 3A

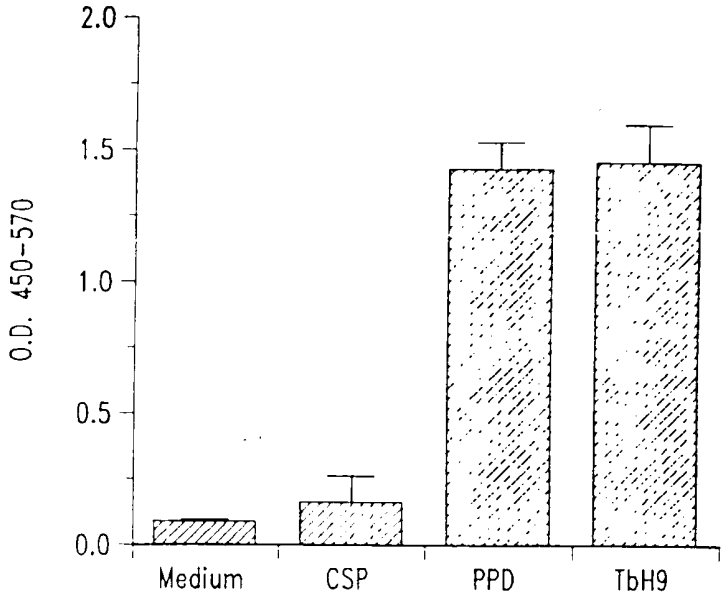


Fig. 3B

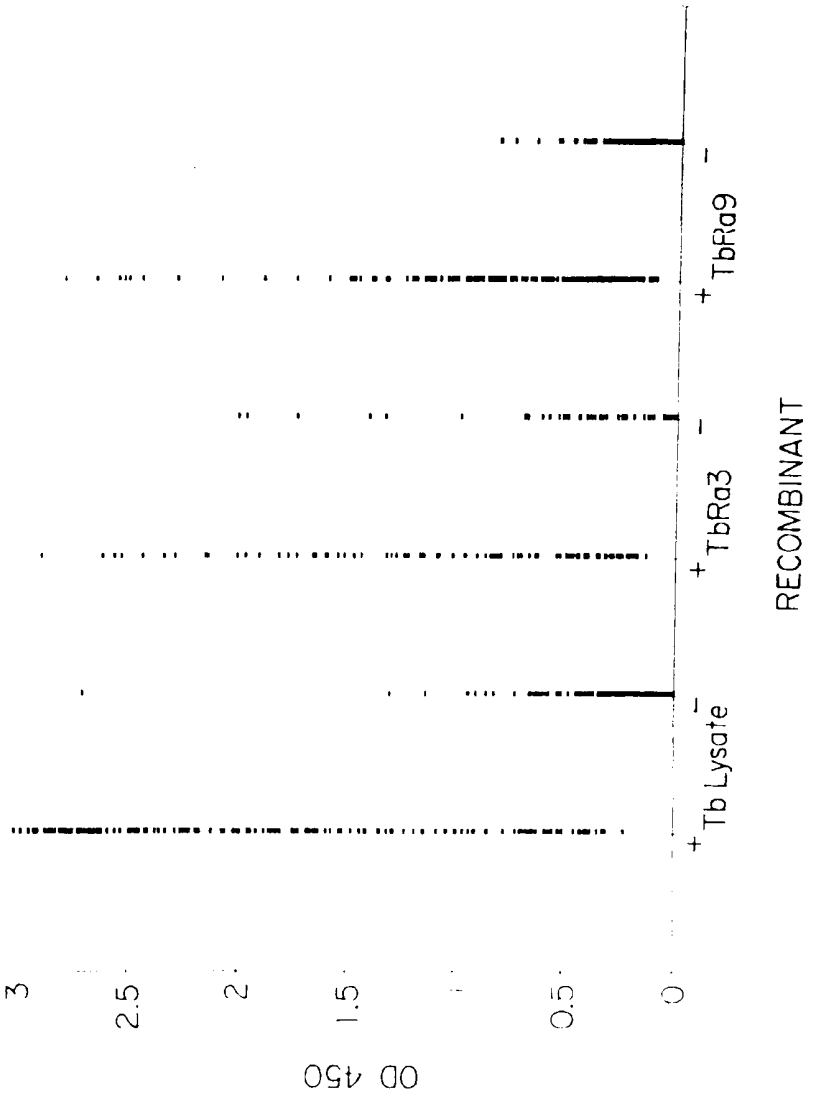


Fig. 4

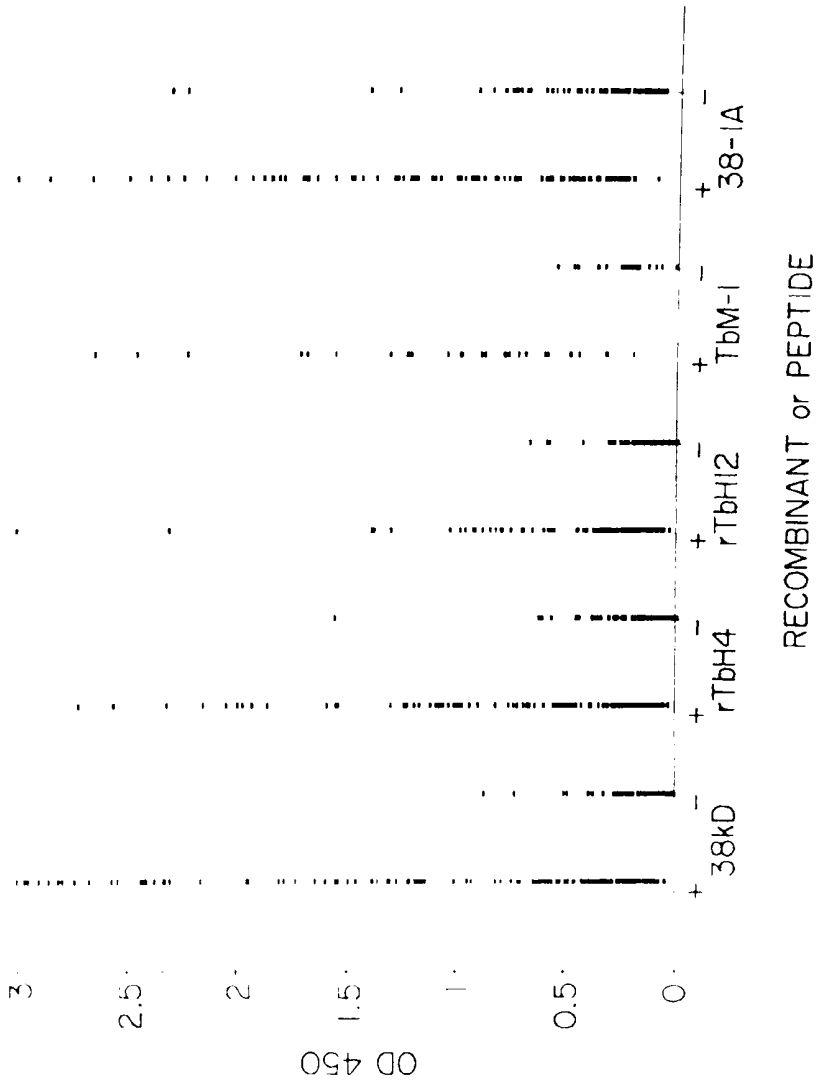

OD 450
3
2.5
2
1.5
1
0.5
0
38kD
rTbH4
rTbH12
TbM-1
38-1A
RECOMBINANT or PEPTIDE

Fig. 5

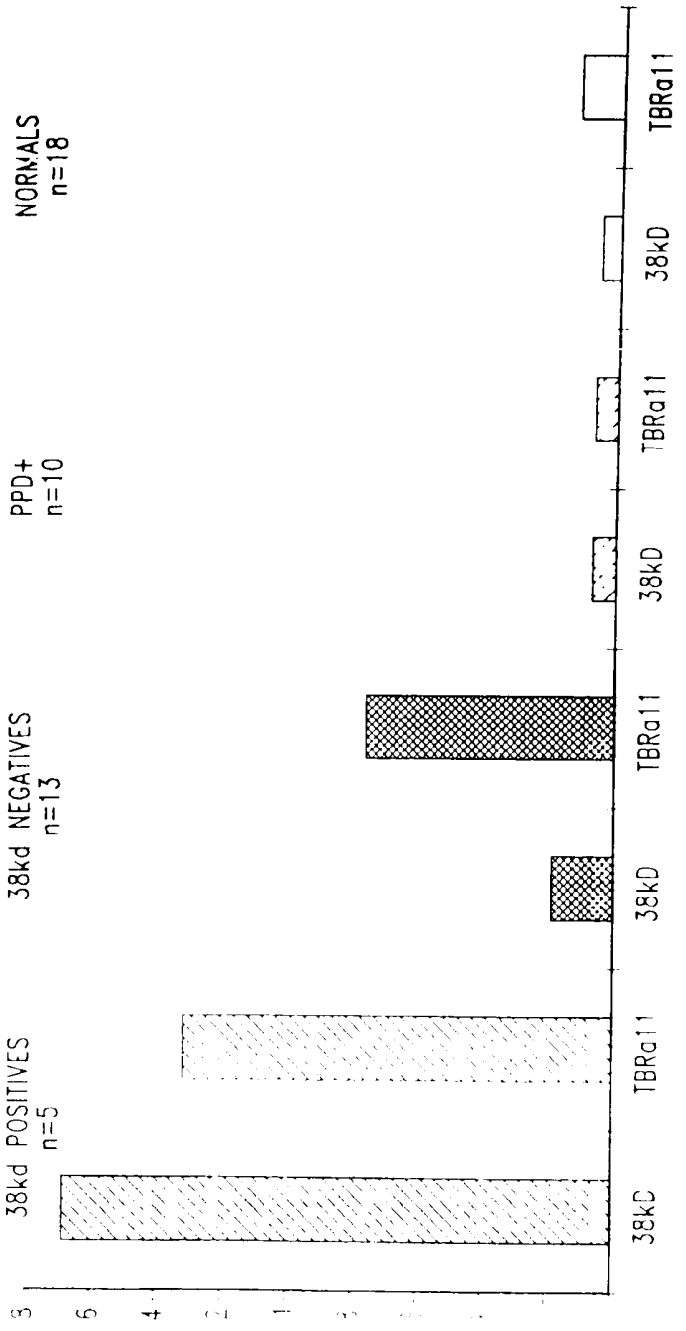

38kd POSITIVES
n=5
38kd NEGATIVES
n=13
PPD+
n=10
NORMALS
n=18
38kD
TBRa11
RECOMBINANT

Fig. 6.

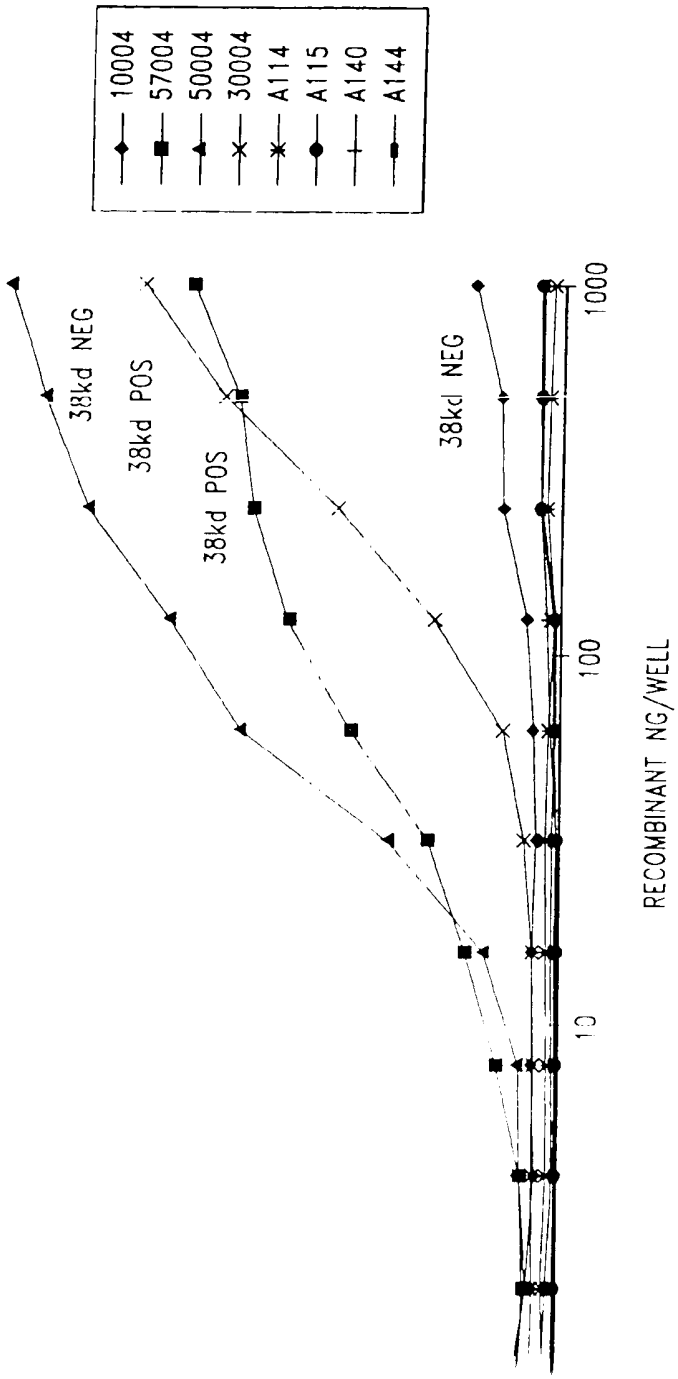

10004
57004
50004
30004
A114
A115
A140
A144
38kd NEG
38kd POS
38kd POS
38kd NEG
1000
100
10
RECOMBINANT NG/WELL

Fig. 7

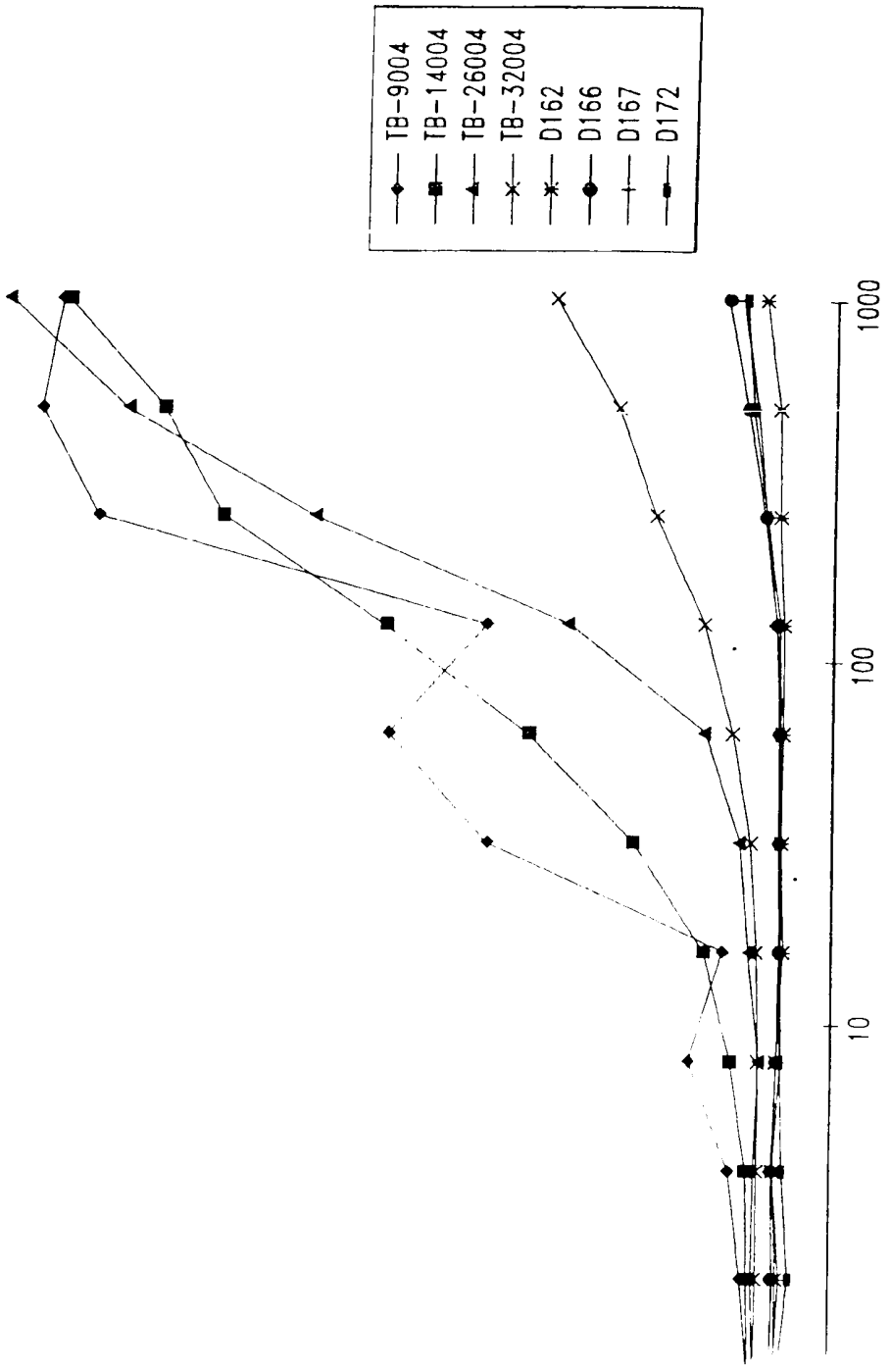

TB-9004
TB-14004
TB-26004
TB-32004
D162
D166
D167
D172
1000
100
10
RECOMBINANT NG/WELL

Fig. 8

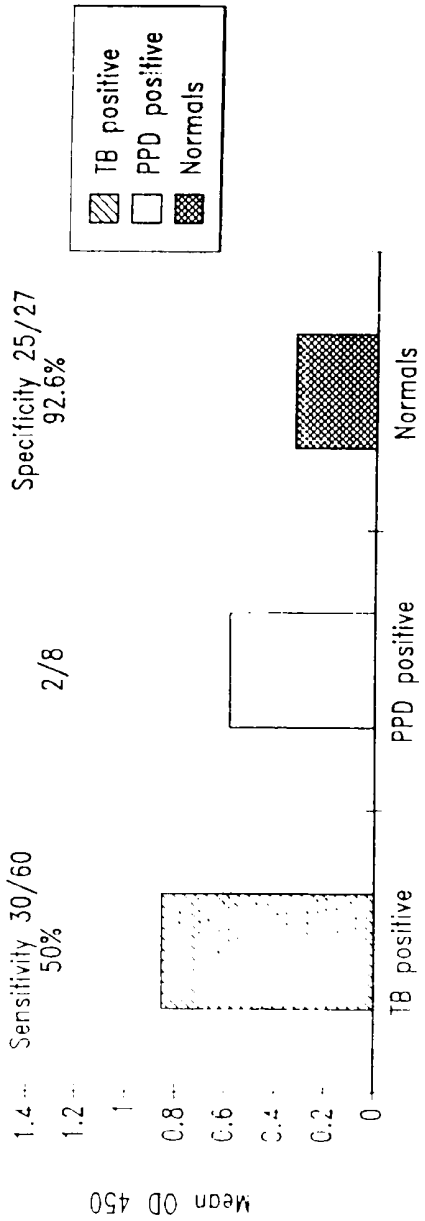

TB positive
PPD positive
Normals
Specificity 25/27
92.6%
2/8
Sensitivity 30/60
50%
1.4
1.2
1
0.8
0.6
0.4
0.2
0
TB positive
PPD positive
Normals
Mean OD 450

Fig. 9

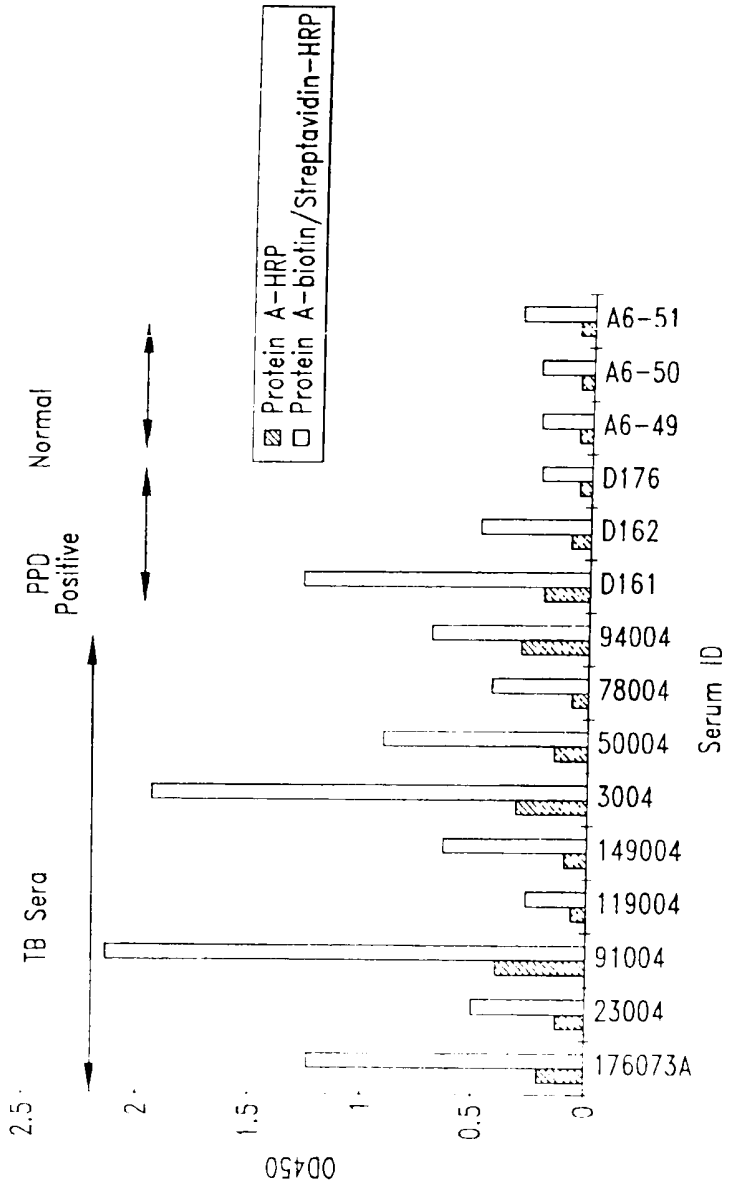

Protein A-HRP
Protein A-biotin/Streptavidin-HRP
Normal
PPD Positive
TB Sera
A6-51
A6-50
A6-49
D176
D162
D161
94004
78004
50004
3004
149004
119004
91004
23004
176073A
Serum ID
2.5
2
1.5
1
0.5
0
OD450

Fig. 10

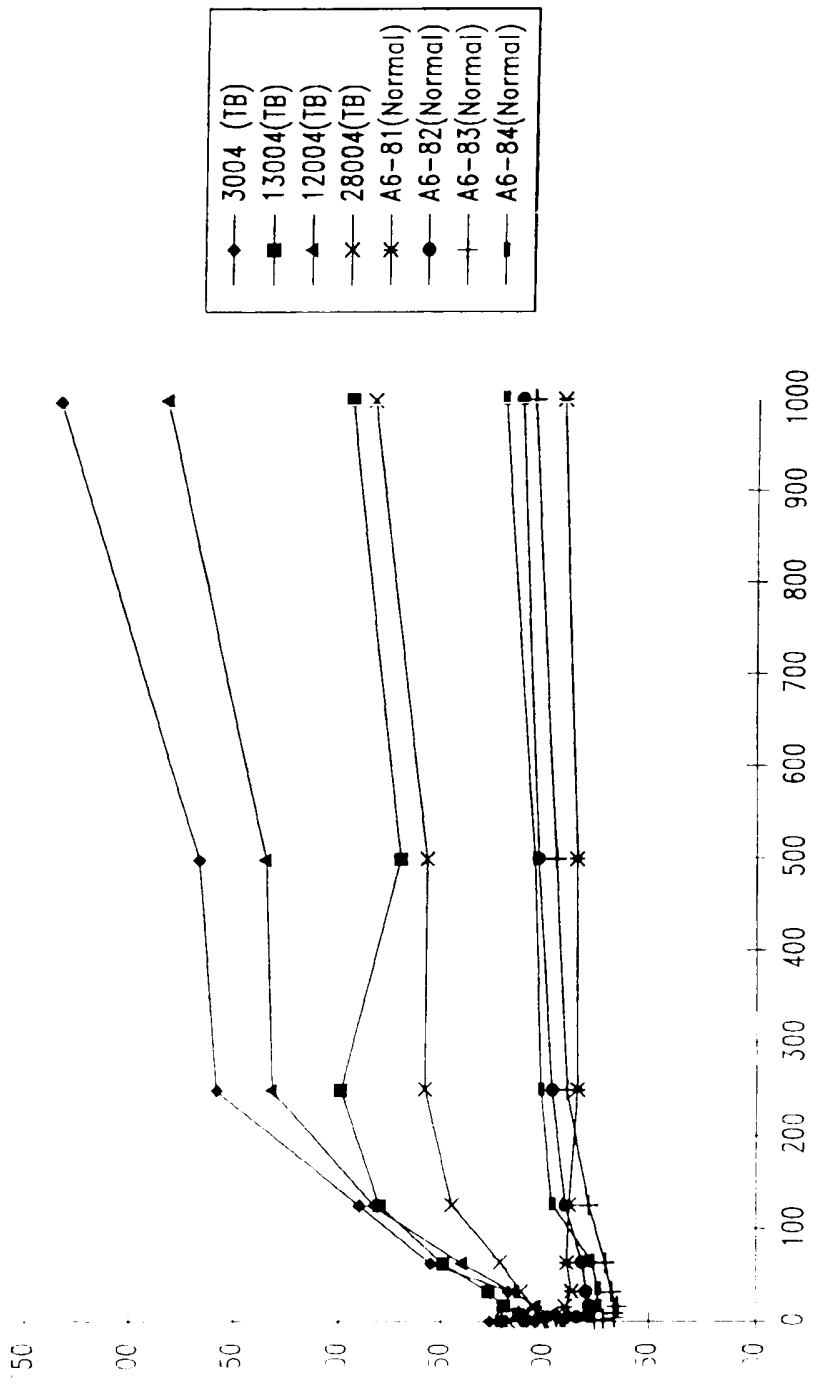

3004 (TB)
13004(TB)
12004(TB)
28004(TB)
A6-81(Normal)
A6-82(Normal)
A6-83(Normal)
A6-84(Normal)
0
100
200
300
400
500
600
700
800
900
1000
Recombinant ng/well

Fig. 11